(Founded in 1925 by Carl Murchison)

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

EDITED BY
CARL MURCHISON

JOHN E. ANDERSON
University of Minnesota

CHARLOTTE BÜHLER
Universität, Wien

CYRIL BURT
University of London

LEONARD CARMICHAEL
Tufts College

SANTE DE SANCTIS
R. Università di Roma

ARNOLD GESELL
Yale University

WILLIAM HEALY
Judge Baker Foundation, Boston

ARTHUR T. JERSILD
Teachers College, Columbia University

BUFORD JOHNSON
The Johns Hopkins University

HAROLD E. JONES
University of California

TRUMAN L. KELLEY
Harvard University

K. S. LASHLEY
Harvard University

A. R. LURIA
Medico-biological Institute, Moscow

HENRI PIÉRON
Université de Paris

Robert R. Sears
State University of Iowa

GEORGE D. STODDARD
University of Illinois

CALVIN P. STONE
Stanford University

LEWIS M. TERMAN
Stanford University

GODFREY THOMSON
University of Edinburgh

E. L. THORNDIKE
Teachers College, Columbia University

C. J. WARDEN
Columbia University

JOHN B. WATSON
New York City

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 39

1949

Copyright, 1949, by The Journal Press
Published quarterly by The Journal Press
Provincetown, Massachusetts, U. S. A.

Table of Contents

No. 1

A STUDY OF THE PSYCHOANALYTIC THEORY OF PSYCHOSEXUAL DEVELOPMENT 3
By Gerald S. Blum

THE ASSESSMENT OF PARENTAL ATTITUDES IN RELATION TO CHILD ADJUSTMENT 103
By Edward Joseph Shoben, Jr.

No. 2

QUALITATIVE DIFFERENCES IN THE VOCABULARY RESPONSES OF NORMALS AND ABNORMALS . . 151
By Herman Feifel

THE RELATIVE EFFECTIVENESS OF MOTION AND STILL PICTURES AS STIMULI FOR ELICITING FANTASY STORIES ABOUT ADOLESCENT-PARENT RELATIONSHIPS 207
By Paul E. Eiserer

THE ORGANIZATION OF HEREDITARY MAZE-BRIGHTNESS AND MAZE-DULLNESS . . . 281
By Lloyd V. Searle

$7.00 per volume
Single numbers $4.00

QUARTERLY
Two volumes per year

February, 1949
Volume 39, First Half

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

Dept. of Psy.

EDITED BY
CARL MURCHISON

John E. Anderson
University of Minnesota

Charlotte Bühler
Universität, Wien

Cyril Burt
University of London

Leonard Carmichael
Tufts College

Sante De Sanctis
R. Università di Roma

Arnold Gesell
Yale University

William Healy
Judge Baker Foundation, Boston

Arthur T. Jersild
Teachers College, Columbia University

Buford Johnson
The Johns Hopkins University

Harold E. Jones
University of California

Truman L. Kelley
Harvard University

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

K. S. Lashley
Harvard University

A. R. Luria
Medico-biological Institute, Moscow

Henri Piéron
Université de Paris

Robert R. Sears
State University of Iowa

George D. Stoddard
University of Illinois

Calvin P. Stone
Stanford University

Lewis M. Terman
Stanford University

Godfrey Thomson
University of Edinburgh

E. L. Thorndike
Teachers College, Columbia University

C. J. Warden
Columbia University

John B. Watson
New York City

Copyright, 1949, by The Journal Press
Provincetown, Massachusetts, U. S. A.

Entered as second-class matter April 15, 1937, at the post-office at Provincetown, Mass., under Act of March 3, 1879

The Journal of Psychology

EDITED BY

Carl Murchison

The Journal of Psychology was founded at the request of many psychologists throughout the world for the purpose of establishing a medium in which publication shall be immediate and in which other publication experiments may be carried out.

(1). The subscription price will be $7 per volume.

(2). The format of each volume will be such that each subscriber may tear down the assembled volume into its component articles.

(3). Except through the fault of the author, the lag in publication should not be more than three or four weeks.

(4). The author will receive 200 reprints of his paper, and subscriber will receive the accumulated publications bound into half volume units. The second half of each volume will contain the title page and table of contents for that volume. Each half volume unit will be mailed to the subscribers at its stated date.

(5). Each author will be charged the actual manufacturing costs of any zinco or halftone cuts.

(6). The publication charges will be as follows:

4	page signature	.....	$22.00
8	" "	.....	32.00
12	" "	.....	48.00
16	" "	.....	64.00
20	" "	.....	80.00
24	" "	.....	96.00
28	" "	.....	112.00
32	" "	.....	128.00
36	" "	.....	144.00
40	" "	.....	160.00

These charges are tentative and will be adjusted as justified by the subscription list (especially by the effect of the war in China and in Europe).

(7). The accumulated units will be mailed to subscribers quarterly: January, April, July, and October. The units may vary in size, to accommodate publication needs.

Two volumes annually. Per annum $14.00; per volume $7.00; per half volume unit $4.00. Complete sets at $7.00 per volume plus transportation.

THE JOURNAL PRESS

2 Commercial Street

Provincetown, Massachusetts

U. S. A.

Single numbers $4.00
$7.00 per volume

QUARTERLY
Two volumes per year

February, 1949
Volume 39, First Half

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

FEBRUARY, 1949

(Manuscripts are printed in the order of final acceptance)

A STUDY OF THE PSYCHOANALYTIC THEORY OF PSYCHOSEXUAL DEVELOPMENT
By Gerald S. Blum

THE ASSESSMENT OF PARENTAL ATTITUDES IN RELATION TO CHILD ADJUSTMENT
By Edward Joseph Shoben, Jr.

Copyright, 1949, by The Journal Press
Provincetown, Massachusetts, U. S. A.
Entered as second-class matter April 15, 1937, at the post-office at Provincetown, Mass., under Act of March 3, 1879.

Published as a separate and in *Genetic Psychology Monographs*, 1949, 39, 3-99.

A STUDY OF THE PSYCHOANALYTIC THEORY OF PSYCHOSEXUAL DEVELOPMENT*[1]

Department of Psychology, Stanford University

GERALD S. BLUM

	Acknowledgments	5
I.	A clinical psychologist views psychoanalytic theory	7
II.	A new approach: The Blacky Test	13
III.	Design of the experiment	23
IV.	From protocol to theory	27
V.	Differences between the sexes	37
VI.	The syntax of psychoanalytic theory	55
VII.	In light of the evidence	71
	Appendices	75
	References	99

*Received in the Editorial Office on October 18, 1948.

[1]A dissertation submitted to the Department of Psychology and the Committee on Graduate Study of Stanford University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, 1948.

ACKNOWLEDGMENTS

I wish to express my gratitude to the many individuals who have made valuable contributions to this study. Foremost are Howard F. Hunt, who served as a stimulating consultant throughout the entire research and subsequent preparation of the manuscript; and Harold L. Raush, who provided an almost daily source of fruitful discussion. I wish to thank Quinn McNemar for his suggestions concerning the design and statistical methodology of the experiment, as well as for his reading of the manuscript. In the latter connection, Maud Merrill James, Myrtle W. Blum, and Lester M. Libo deserve special mention. The artist, whose talented drawings made the study possible, is Alfred S. Arnott. To the host of others who must remain unnamed, I can only state my deep appreciation. Finally, I wish to acknowledge my indebtedness to the Palo Alto Veterans Administration Hospital, which was the scene of much of the preliminary research and later analyses of the data.

GERALD S. BLUM

Department of Psychology
University of Michigan
Ann Arbor, Michigan

I. A CLINICAL PSYCHOLOGIST VIEWS PSYCHOANALYTIC THEORY

Academic psychologists have long been concerned with the issues raised by psychoanalytic theory. Their search for answers has run the gamut from laboratory exploitations of rats, cats, dogs, and sheep to their own introspections on the experience of being psychoanalyzed. In general, these endeavors have been directed toward determining whether psychoanalytic theory can sufficiently withstand the rigor of scrutiny by psychologists so as to qualify for membership in the subtle society known as Science.

For the emerging clinical psychologist, however, this problem has largely been by-passed. When he looks about him, he discovers that the principles of psychoanalytic theory are generally the vogue. No longer is it a question of whether the evidence is sound enough to be admissible in court. The jury now consists of 12 people named Freud. Consequently the neophyte sets about to verse himself in these new ways and, as his clinical experience broadens, he tends to find that the psychoanalytic framework applies fairly well to the cases he has seen. Before long, his Rorschach interpretations come to acquire the authentic flavor of Old Vienna and psychoanalytic theory gains a place in the desk drawer alongside of the 10 inkblots.

It is at this stage that the young clinical psychologist conceivably can become perturbed over his recent awakening to the fact that there is not one world. He finds himself in the process of shedding the academic mantle of skepticism and wonders if his new-found clinical insight will be enough to keep him warm. Is he willing to have applied to himself the stereotype: "clinician"—a person who may be quite good at weaving "dynamic" personality patterns but who has largely forsaken the realm of pure truth? Assume that, because of certain factors (clinically ascertainable, of course) in his own personality, he is unwilling to make the choice. Suppose, instead, that he wants to try to reconcile his two worlds. What data can he call upon? What knowledge, acceptable to both academician and clinician, relevant to psychoanalytic theory currently exists?

Before seeking answers to these questions, it might be well to review very briefly the portion of psychoanalytic theory with which this study purports to deal: the theory of psychosexual development. The latter is of fundamental importance in psychoanalysis, for stages of psychosexual development from birth through puberty have been delineated, and the

course of an individual's progress through these stages is considered to be of major significance in the shaping of his personality. The sequence of stages is a fixed one, though all phases gradually pass into one another and overlap.

At birth the infant enters the oral sucking stage, during which he is concerned primarily with the pleasurable autoerotic stimulation of the oral erogenous zone. External objects play no part at this time. Next, he proceeds to the oral sadistic phase, in which he can recognize an object, but his aim is to incorporate the object in a destructive fashion. In the second year of life, the anal erogenous zone comes to the fore, and in this period the infant first derives pleasure from defecation and later comes to realize the sadistic power which he can exert over his parents by giving or retaining the feces. The fourth and fifth years furnish the scene for the phallic or early genital phase, in which major developments are the Oedipus complex, masturbation, fear of castration in boys, and penis envy in girls. From the sixth or seventh year until puberty the latency period follows, when psychosexual development appears to assume a rôle secondary to social growth. Finally, in puberty and afterward, the late genital stage is expressed in so-called "normal" sexual adjustment.

Thus, adult personality is considered to be dependent in a large measure upon the manner in which these psychosexual stages were experienced in childhood. In this connection, the concepts of fixation and regression are important. Progress to a higher stage never takes place completely, for characteristics of the lower level persist to some extent. Disturbances may result in an arresting of development at any stage, which is termed fixation, or they may cause the retention of an abnormal number of characteristics of an earlier stage, to which the individual will return if difficulties arise — the latter process being known as regression.

Accompanying the advance through this genetic sequence are changes in the types of object relationships established during psychosexual development. At first the infant is unable to differentiate objects, but after this capacity has gradually been acquired, the mother becomes the first object of every individual — due to the fact that it is the mother who cares for its needs. This early mother attachment makes the development of object relationships simpler in the case of the boy, for in the later oedipal situation the mother remains as his primary object choice. In the Oedipus complex, which is described as the climax of infantile sexuality, the child is more strongly attracted to the parent of the opposite sex and is jealous of the parent of the same sex. This involves a complicating transition in the girl's

love from the mother to the father — a transition which presumably arises, between the ages of three and six, from disappointment over the lack of a penis, for which the mother is held responsible. Consequently the girl turns away from the mother to the father. In the male the Oedipus complex is given up as a result of castration anxiety, for the boy fears that the hostile father will punish his sexual strivings toward the mother by castrating him. No comparably dramatic process occurs in the female.

The resolution of the Oedipus complex involves the concepts of identification and superego. Identification is a regressive type of object relationship in which the individual, who is frustrated in his attempts to attain his love object, resorts to incorporation within himself of the characteristics of his successful competitor. For example, the boy who is frustrated in his desires for the mother proceeds to pattern himself after his rival, the father. Similarly the girl tends to take after the mother and hence enjoys a sort of vicarious satisfaction through the father's love for the mother. The superego arises from a process whereby parental prohibitions are internalized or introjected. The child learns to incorporate the threats of his parents, so that a portion of his own personality assumes the function of controlling his actions. The model for this introjection is most frequently a mixture of both parents, but predominantly the parent who is regarded as the source of the most decisive frustrations. In addition to the negative, punishing aspect of the superego, there is the positive side or ego ideal.[2] The ego ideal develops somewhat later than the superego and involves the incorporation by the child of his parents' positive ideals and good wishes.

Other important object relationships are sibling rivalry and the type of love-object choice. Sibling rivalry may be due to a variety of motivating factors; for example, the child may resent a competitor for the parents' affection; or he may repress his hostility toward the parents and displace it on to a sibling; or he may express aggressions arising from a general feeling of insecurity. But whatever its etiology may be, it is considered to be of major significance in personality development. As for selection of a love-object, two types have been distinguished: the anaclitic type of choice — in which an object is chosen because it provokes associations about an original object of the past, usually the parent of the opposite sex — and the narcissistic type of choice, in which an object is chosen because it represents some characteristic of the individual's own personality.

The foregoing presents, of course, a highly oversimplified account of the

[2]The controversy over whether the positive ego ideal can legitimately be distinguished from the superego is not crucial to the present study.

psychoanalytic theory of psychosexual development. It is merely intended to serve as a framework within which to consider the available evidence on the question in point: What does existing research have to offer a clinical psychologist who desires answers concerning the scientific accuracy of psychoanalytic theory?

Some years ago Sears (12) prepared a survey of such studies. One of the studies on oral eroticism reported is that done with puppies by Levy. Levy in 1934 compared the frequency of body-sucking and the sucking of a proffered finger by puppies having different amounts of opportunity for sucking. A litter of six animals was taken at birth. One pair, the long-feeders, were fed by bottle with a small hole in the nipple and after each meal were allowed to suck to satiety on a nipple-covered finger. During the 20 days of the experiment they averaged 62 and 67 minutes of sucking per 24 hours. The short-feeders were fed with a large-holed nipple and were not given a finger to suck afterward. They averaged 21 and 22 minutes of sucking per 24 hours. The third pair from the litter, the breast-feeders, were left with the mother. The short-feeders gave evidence of sucking deprivation by chewing and sucking at each other's bodies between meals, and by being far more responsive to a proffered finger between meals than were the long-feeders. The long-feeders were responsive to the finger just before meals but not at other times. The breast-feeders were never responsive to the finger and did no body-sucking. Sears summarizes a survey of several studies on oral eroticism with the comment: "[Evidence] points rather definitely to Freud's contention that pleasure sucking is detached early in the infant's life and henceforth can provide gratification independent of that provided by nutritional sucking or the ingestion of food" (12, p. 8).

With respect to castration anxiety and penis envy, he reports a survey conducted by Hattendorf in 1932. Hattendorf secured information from mothers about the sex questions their children had asked. A corps of social workers and other trained investigators called on mothers in Minneapolis homes and obtained the data by direct questioning. The list of questions shows that there was a very widespread recognition by children of both sexes that there was a physical difference in the genitalia. But only three of the 137 questions asked by children 2 to 5 years of age suggested that boys thought the girl's lack of a penis was the result of injury. There is no evidence, either, that girls envied the penis or wished to be boys. Sears states: "The data obviously do not disprove the possibility of such reactions as Freud has described but their universality is not only not demonstrable, but is flatly disproven" (12, p. 30). He concludes further that Freud seriously

overestimated the frequency of the castration complex, which is a "function of the kinds of information children have" (12, p. 36).

Concerning masturbation, Sears refers to a summary of the literature made by Willoughby in 1937, in which only 5 per cent of men and 18 per cent of women could recall having masturbated before 10 years of age. In another study, Levy had reported that, on questioning the mothers of 49 boys, genital manipulation had been observed in 53 per cent by the age of three years. Sears reconciles the two divergent findings by saying that masturbation was probably not defined similarly.

Turning next to the Oedipus complex, there is an account of a study by Terman in 1938, in which he had his subjects in the marriage study rate, on a five-point scale, both attachment to and conflict with each parent. Sears summarizes the results as follows:

> Contrary to the Freudian theory, Terman found that there was little or no difference between the sexes in the amount of attachment to each parent, and that in both cases it was greater with the mother. The ratings on amount of interpersonal conflict indicate somewhat less between boys and their mothers than between boys and their fathers; there is a very slight tendency of the opposite sort with girls. In all cases the differences are too small to be of great import. Taking the population as a whole, there is no support for the theory that the cross-sex parent is favored and that powerful jealousy reactions are developed toward the same-sex parent (12, p. 42).

Later Sears adds that the Oedipus relationship is too dependent on specific conditions of learning to be demonstrable in large populations.

In summarizing his impressions from a number of such "objective studies," Sears concludes that other social and psychological sciences must gain as many hypotheses and intuitions as possible from psychoanalysis but that the further analysis of psychoanalytic concepts by non-psychoanalytic techniques may be relatively fruitless so long as those concepts rest in the theoretical framework of psychoanalysis. Instead he pleads for the development of a science of personality that is behavioral in character.

Is this, then, the answer which a clinical psychologist seeks? Will he now admit that he and many of his confreres, while *possibly* not barking up the wrong tree, have nevertheless been barking much too loudly? Or will he instead look rather askance at such conclusions and reserve his judgment on the grounds that the psychoanalytic theory which he knows has not actually been under investigation in many of the studies reported? Perhaps he can agree with Sears that further research along the same lines may be

relatively fruitless, but whether the fruitlessness lies in the theoretical framework of psychoanalysis or in the theoretical framework of the non-psychoanalytic techniques used to study it, remains in doubt. Certainly a clinician would not expect an experiment in which subjects are asked to rate attachment to parents to be crucially pertinent to the validity of the Oedipus complex. If the theoretical framework of psychoanalysis is based upon such subtle concepts as repression, reaction formation, denial, and the like, it would appear irrelevant to attempt an appraisal of the theory by a technique which makes no allowance for these very concepts!

So the general question of the scientific acceptability of psychoanalytic theory does not seem to have a satisfactory, ready answer. This might well be a source of concern to all research-minded psychologists. But to the clinician its significance is more special, for he could not be content even with evidence which would give general support to the theory which he utilizes in practice. He would want to know, not merely that psychoanalytic theory as a whole seems to have a fair degree of merit, but, more specifically, which aspects of the theory possess demonstrable validity and which do not. He would want to evaluate the theory in all its complex interrelationships, for these are the things with which he deals in his work.

To be sure, in light of the available data this would seem to be asking a ot. But perhaps a start can be made in the right direction. A clinical sychologist presumably understands psychoanalytic theory and its implica-ions for the study of personality. He also has at his disposal the latest nowledge about psychological techniques for investigating personality. he task becomes one of combining these assets in the most fruitful way. he experiment to be described represents such an attempt.

II. A NEW APPROACH: THE BLACKY TEST

In casting about for an appropriate psychological technique to use in studying the psychoanalytic theory of psychosexual development, the area of projective personality testing seemed to offer the most promise. According to this method, an individual, when confronted with a fairly ambiguous situation, identifies himself with a figure in the situation and ascribes his own unconscious reactions to that figure. Thus, the deeper, underlying aspects of personality are tapped in a manner which is not feasible with less subtle instruments. In this respect a projective technique has some a priori relevance to psychoanalytic theory, for the latter also deals largely with unconscious processes.

Once limited to the field of projective testing, the task then centered upon which instrument to use for the purposes of the experiment. The two more or less standard projective techniques in current clinical practice are the Rorschach Test and the Thematic Apperception Test.[8] Personality interpretations based upon a subject's performance on these tests are frequently psychoanalytically oriented. There are numerous test clues which can be fitted into the psychoanalytic context. But such interpretations are largely indirect and involve a great many inferences from the data. Therefore it was felt that a new approach, utilizing advantages of the projective method while at the same time related more directly to psychoanalytic theory, might be more fruitful.

To meet this need, the author devised a technique called the Blacky Test. It consists of 12 cartoon drawings or caricatures (Figures 1-12), which depict the adventures of a dog named Blacky. The first cartoon is an introduction to the cast of characters, which includes Blacky, Mama, Papa, and Tippy (a sibling figure of unspecified age and sex). Each of the 11 subsequent cartoons is designed to portray either a stage of psychosexual development or a type of object relationship within that development:

Cartoon I (Fig. 2)	Oral Eroticism
" II (Fig. 3)	Oral Sadism
" III (Fig. 4)	Anal Sadism
" IV (Fig. 5)	Oedipal Intensity
" V (Fig. 6)	Masturbation Guilt
" VI (Fig. 7)	Castration Anxiety (Males) Penis Envy (Females)

[8]It is assumed that the reader is familiar with both of these tests. If not, see Klopfer, B., & Kelley, D. McG. *The Rorschach Technique.* New York: World Book Company, 1942; and Tompkins, SS. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.

	"	VII (Fig. 8)	Positive Identification
	"	VIII (Fig. 9)	Sibling Rivalry
	"	IX (Fig. 10)	Guilt Feelings
(Males)	"	X (Fig. 11)	Positive Ego Ideal (Males)
(Females)	"	XI (Fig. 11)	Love-Object (Females)
(Males)	"	XI (Fig. 12)	Love-Object (Males)
(Females)	"	X (Fig. 12)	Positive Ego Ideal (Females)

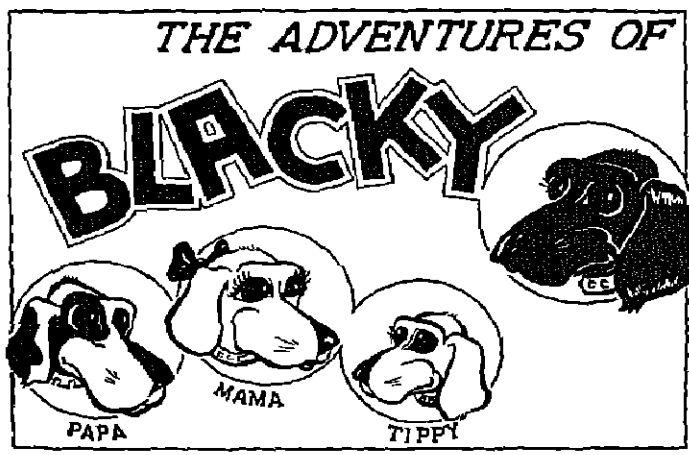


FIGURE 1
CAST OF CHARACTERS

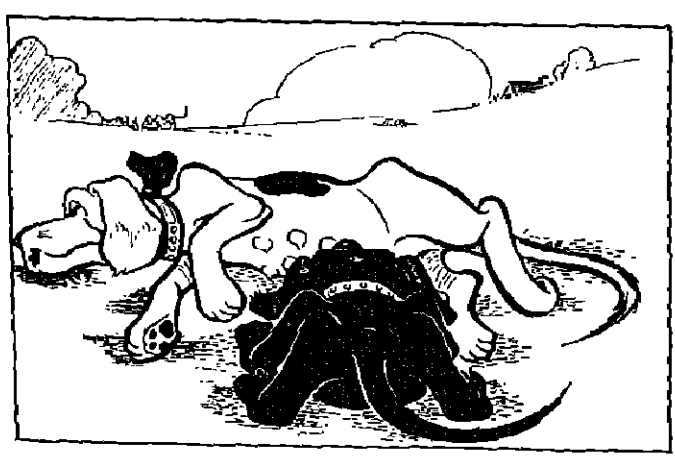


FIGURE 2
CARTOON I: ORAL EROTICISM

The same cartoons are used for both male and female subjects, but the sexes are tested separately. When presented to males, Blacky is described as the "son" and the cartoons are shown in the above order. When presented to females, Blacky is described as the "daughter" and Figure 12 precedes Figure 11.

Before going further into the technique itself, it might be well to antici-

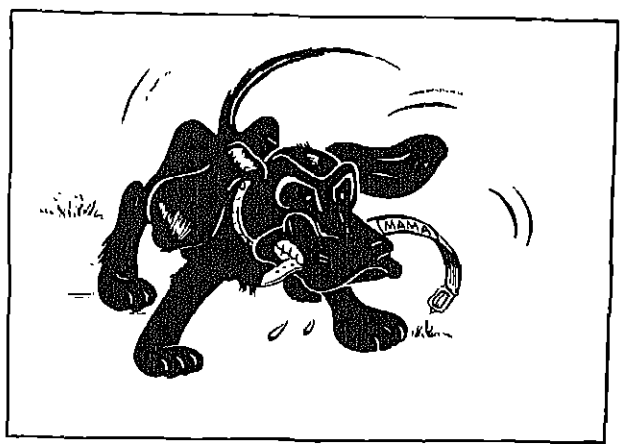


FIGURE 3
CARTOON II: ORAL SADISM

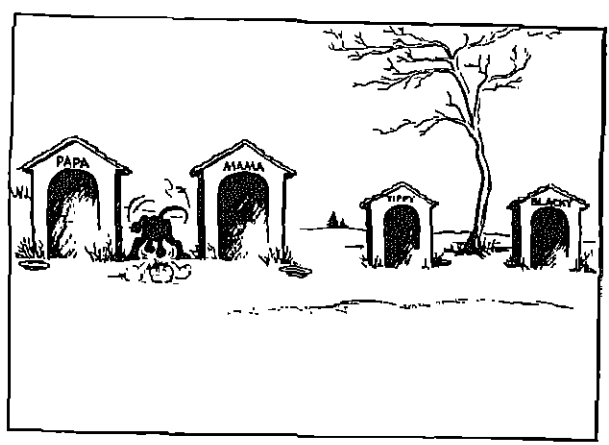


FIGURE 4
CARTOON III: ANAL SADISM

pate some of the reader's questions at this point: Why dogs? Why the names "Blacky" and "Tippy"? The characters were relegated to dogdom in order to facilitate freedom of personal expression in situations where human figures might provoke an unduly inhibiting resistance — in other words, might be "too close to home." While on the one hand minimizing the dangers of resistance, the canine medium, thanks to Disney cartoons and comic

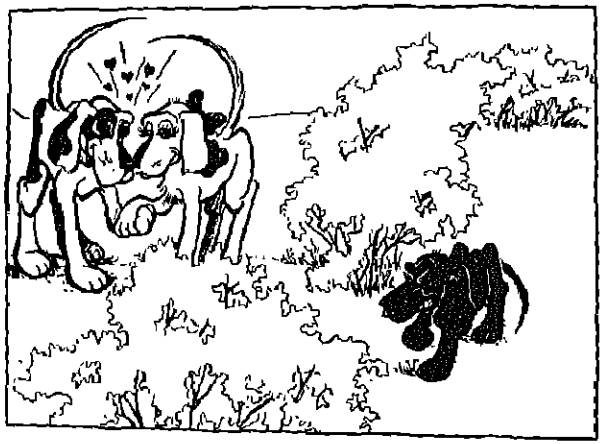


FIGURE 5
CARTOON IV: OEDIPAL INTENSITY

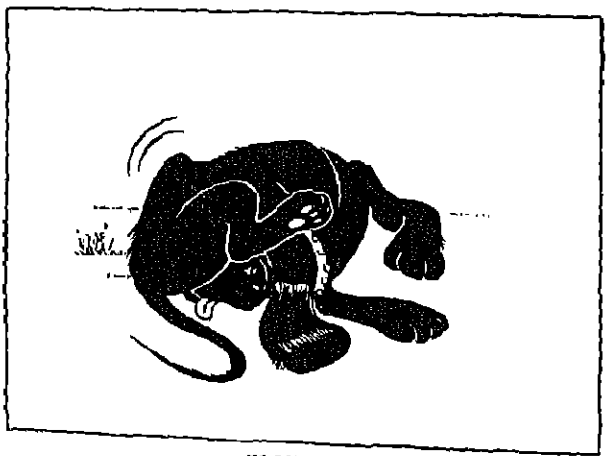


FIGURE 6
CARTOON V: MASTURBATION GUILT

strips, still preserves sufficient reality so that subjects can identify themselves quite fully with the cartoon figures and project their innermost feelings. The ease with which people "humanize" a dog's actions will be readily apparent from sample protocols which will be given later.

The names "Blacky" and "Tippy" were selected by means of an informal survey in which a list of approximately eight dog names was shown to a

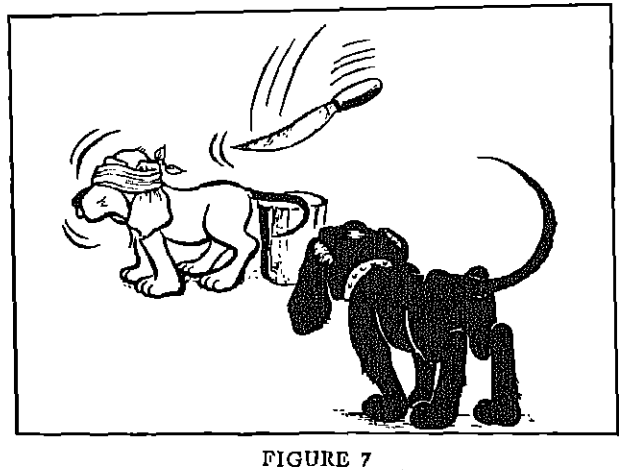


FIGURE 7
CARTOON VI: CASTRATION ANXIETY (MALES); PENIS ENVY (FEMALES)

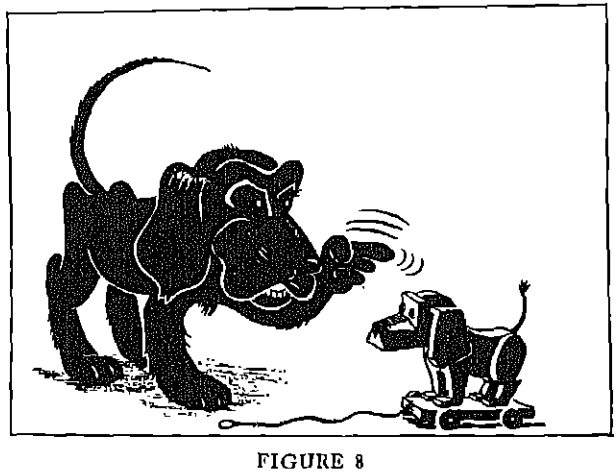


FIGURE 8
CARTOON VII: POSITIVE IDENTIFICATION

small group of males and to a small group of females. All the persons were asked to record whether they felt that each name suggested the male sex or the female sex. Generally the male judges felt "Blacky" to be male, whereas the female judges regarded it as female. "Tippy," who was intended to play the rôle of either a brother or sister for both male and female

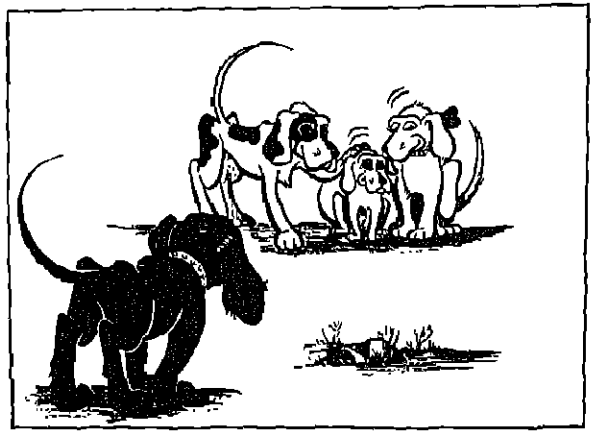


FIGURE 9
CARTOON VIII: SIBLING RIVALRY

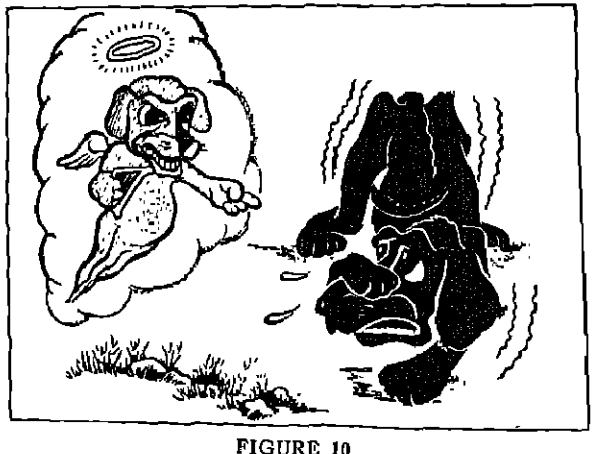


FIGURE 10
CARTOON IX: GUILT FEELINGS

subjects, had his (her) sex described equally often as male or female by both male and female judges. The christening took place shortly thereafter.

The test, when administered in the *group* form,[4] is projected onto a screen by means of $3\frac{1}{4}'' \times 4''$ lantern slides. The following preliminary instructions are given:

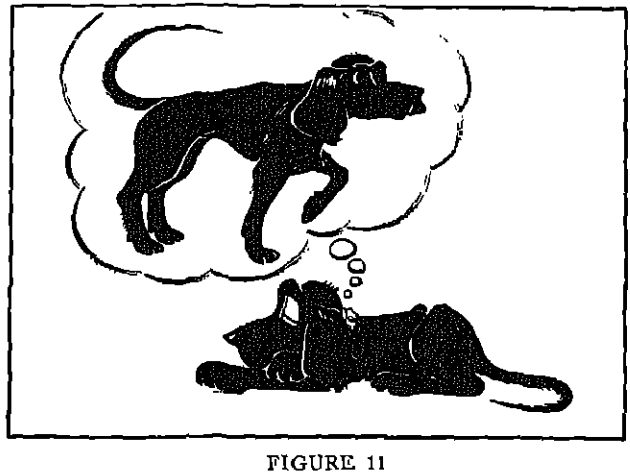


FIGURE 11
CARTOON X: POSITIVE EGO IDEAL (MALES); LOVE OBJECT (FEMALES)

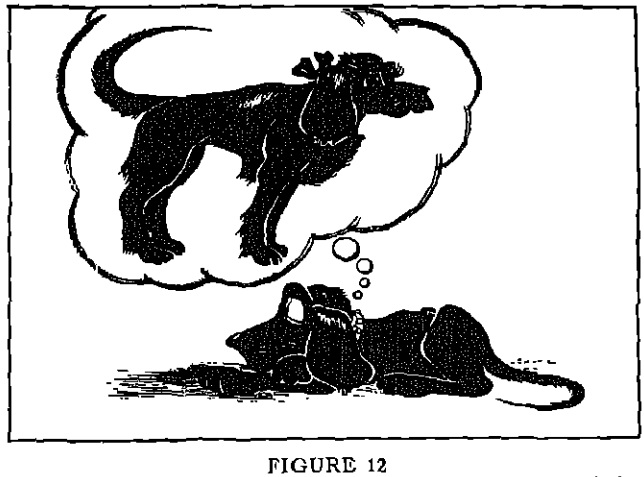


FIGURE 12
CARTOON XI: POSITIVE EGO IDEAL (FEMALES); LOVE OBJECT (MALES)

[4]In clinical practice the test is usually administered individually.

> What we have here is a bunch of cartoons, like you see in the funny papers, except that there are no words. We'll show them to you one cartoon at a time and the idea is for you to make up a little story about each one—just tell what is happening in the picture, why it's happening, and so on. Since this is sort of a test of how good your imagination can be, try to write vividly about how the characters feel. You will have two minutes for each story, which means about one or two paragraphs on each cartoon. It is desirable to write as much as possible within the time limit.
>
> At the end of each cartoon there will be some questions about them, which are to be answered below the stories. There are no right or wrong answers to these questions—just what you imagine the answer to be. For each question pick the one answer which seems to fit best. Don't bother about such things as grammar or spelling—we're only interested in the content of your stories.
>
> Before we start, here are the characters you'll see in the cartoons. (Figure 1 is shown for about 20 seconds.) There's Papa, Mama, Tippy, and the son (daughter), Blacky, who is the main figure in the stories. Now for Cartoon I. . . .

Several comments concerning these preliminary instructions are in order. The general flavor of encouraging vivid, dramatic, imaginative stories is similar to the directions accompanying the Thematic Apperception Test. The two-minute time limit was eventually settled upon as the most suitable compromise between the goals of sufficient story-length to bring out significant material and sufficient time-pressure to facilitate spontaneous rather than premeditated expression. Blacky's sex was purposely specified to make it easier for the subject to identify with the cartoon figure. The sequence of administration — spontaneous story followed immediately by questions in the inquiry[6] — was adopted in order to maintain the original "set" of the subject rather than to have his responses be a function of a perspective gleaned from the total test.

The questions in the Inquiry (Appendix A) are mostly of the multiple-choice type, in which the subject is asked to pick the single alternative which he considers to be most applicable, plus some direct questions requiring one or two sentences to answer. The significance of the individual alternatives was assigned in the same a priori manner as in the case of the cartoons. Six or seven questions accompany each cartoon. The Inquiry currently in use was developed over a period of about a year, during which time tentative forms of the test were administered to several groups of elementary

[6] This differs from the usual Rorschach Test procedure of conducting the inquiry after all 10 cards have been presented.

psychology students at Stanford University. Revisions were made from inspection of the item-counts, those questions being retained which seemed to distribute the population satisfactorily.

After the preliminary directions have been given and the cast of characters has been presented, Cartoon I, introduced with the comment "Here is Blacky with Mama," is shown. The group of subjects is allowed two minutes (a "30-seconds to go" announcement is made) to write its stories, and then each question is flashed on the screen for several seconds. A similar procedure ensues for the remaining 10 cartoons, each accompanied by an introductory comment. Of the latter, nine comments are purely descriptive and do not serve to structure the situation beyond what is depicted in the cartoon. On Cartoons III (Anal Sadism) and V (Masturbation Guilt), however, it was found that the situations were often misinterpreted and some additional structuring was deemed desirable. This was supplied in the statements "Here Blacky is relieving himself (herself)" on III and "Here Blacky is discovering sex" on V.

Upon finishing the inquiry on Cartoon XI, the subjects are told:

> Draw a line beneath what you have just finished and write down the numbers 1 through 11. We will show each cartoon again for only a few seconds and you are to record opposite its number whether you like or dislike the cartoon. Use *L* for Like and *D* for Dislike. . . . Now, from the ones you like, pick the single one which you like best and record its number opposite the word "Best." If you want to have another look at all the cartoons together, they have been placed in order along the blackboard. Then write down your reasons for selecting that particular cartoon as best. . . . Now, from the ones you dislike, pick the single one which you dislike the most and record its number opposite the word "Worst." Again you may refer to the cartoons along the blackboard. Then write down your reasons for selecting the one which you picked as worst.

The test is concluded with the request that each subject record what he considers to be the purpose of the test, plus any other reactions, destructive as well as constructive, which he cares to express. He is also asked to furnish the following information about his siblings:

1. Siblings listed as "Male" or "Female" in chronological order, including himself in the sequence.
2. Sibling(s) living or not.
3. Approximate age of sibling(s) if alive or approximate date of death if not.

By now the reader may be getting justifiably impatient over how all this ties in with an attempt to study psychoanalytic theory. But first one further digression. The purist in clinical psychology might be offended by the inclusion of the Blacky Test among "projective techniques." He might argue that the situations are too highly structured to fit the literal definition. The arbitrary distinction in definition is of no concern in this study. The test consists of a set of stimuli designed to elicit responses from a group of subjects. The responses in the test situation are personal samples of behavior—"projected" in the sense that qualities are attributed to the stimuli. They differ from subject to subject and consequently are legitimate data for investigation. For the sake of the purist, however, the Blacky Test will henceforth be referred to as a "modified projective technique."

III. DESIGN OF THE EXPERIMENT

The logic involved in the design of this experiment requires close scrutiny. The Blacky Test, a modified projective technique, is being used as a medium to study the psychoanalytic theory of psychosexual development. What are the assumptions? The hypotheses? What conclusions may be drawn?

First there is the assumption that the Blacky Test is actually measuring the psychoanalytic dimensions which it is intended to measure. Apart from the face validity of the test, seconded by the few psychoanalysts to whom it has been shown, the only evidence currently available comes from informal clinical support of test findings on a number of mental hospital patients. The latter source has not as yet been systematically explored and therefore the validity of the test is still indeterminate.

Assume for the present that the test is making valid discriminations. The following hypothesis can then be set up: "If the postulates of psychoanalytic theory are correct, results obtained with the Blacky Test should be consonant with those postulates." The investigation of this hypothesis can be pursued along these two lines: (*a*) sex differences in psychosexual development, as revealed by the test; and (*b*) interrelationships of those psychoanalytic dimensions considered by the test. For example, if males and females differ significantly in degree of narcissism as measured by the Blacky Test, is this difference in accord with psychoanalytic theory or not? Or, if on the test a positive correlation is found between anal expulsiveness and castration anxiety, does this finding agree or disagree with the theory?

One may concede the ideal merit of the above scheme but still question the value of a procedure which may in reality be based on a gratuitous assumption. What meaning can legitimately be attached to results under the existing condition of uncertain test validity? Here the reader must be cautioned to follow the thread of the argument carefully. Every test result (sex differences and dimensional intercorrelations) will show either a statistically significant or insignificant relationship. Both significant and insignificant test findings are types of evidence. However, the two types do not lend themselves equally to interpretation. Insignificant test findings, whether they support the theory or not, must remain enigmatic, for it cannot be known if the absence of relationship is due to a true absence or to an artificial absence produced by failure of the test to measure what it is intended to measure. Significant test findings, on the other hand, can be interpreted *more unequivocally,* for in such cases the test appears to be *measuring something.* Consideration of results in this *exploratory* research

will therefore be confined to the statistically significant.[6]

There are several possible ways to account for results which show statistical significance in the experiment. One possibility is that they may be due to chance. In a large number of statistical comparisons it is expected that a certain percentage will reach statistical significance on a chance basis alone. If the actual number of significant findings does not exceed this chance expectation, the search for explanations can stop in its tracks. But if the actual number is sufficiently greater than would be predicted from the operation of chance factors alone, the search must be resumed. A second possibility is explanation in terms of artifacts within the test. A positive correlation, for instance, may be attributable to hidden features in the construction of the test rather than to real agreement on the dimensions being investigated. Artifacts of this sort can be identified through careful examination of the instrument.

Statistically significant test findings, which are not accounted for by chance or artifact, can then shed some light of their own. Are they consistent with psychoanalytic theory? Affirmative answers *lend support* to the theory. Negative answers *cast some doubt* on the theory. The issue of definitive proof or disproof cannot arise because of the tentative validity of the experimental test. But the answers can be *strongly suggestive* in formulating an independent evaluation of psychoanalytic theory.

The criterion against which psychoanalytic theory is to be compared, therefore, is significant data on the Blacky Test. Affirmative results (favorable to the theory) are those cases in which the theory is found to agree with the significant test findings. Negative results (unfavorable to the theory) are those cases in which the theory is found to disagree with the significant test findings. Obviously the a priori decision to restrict the study to discussion of statistically significant test data in no way biases subsequent agreement or disagreement with the theory. Insignificant findings can be consistent with psychoanalytic theory or not; significant findings can be consistent with psychoanalytic theory or not. The study is restricted to the latter because attempts to account for the former are more conjectural. But some very definite limitations are introduced by this restriction, for the criterion is incomplete (insignificant test findings are excluded from consideration) and the sampling of theoretical postulates might very well be unrepresentative.[7] In view of these limitations, the

[6] The reader who wishes to examine insignificant test findings will be given the opportunity to do so. These test results, while not treated specifically in the text for the reasons cited above, will be presented in the Appendix.

[7] Restricting the study to significant test data selects, in an unknown fashion, those theoretical postulates which will come under consideration.

study must be regarded as an *exploratory first step* which seems logical and worth while.

Before describing the actual procedures of the experiment, it should be added parenthetically that psychoanalytic theory is the only one being investigated. The study carries no implications for any other theoretical position.

In accordance with the foregoing rationale, the Blacky Test was administered to male and female subjects drawn from a population of elementary psychology students at Stanford University during the Winter and Spring quarters of 1948. The subjects "volunteered" in a rather restricted sense of the word. Each student was required to serve a specified number of hours as a research subject, with the choice of research projects largely up to the student himself. All of the male subjects were tested during the Winter quarter, but the disproportionate sex ratio necessitated carrying the testing of females over into the Spring quarter. A large majority of the enrolled female students participated in the experiment, and considerably more than half of the enrolled males took part. All of the subjects were tested early in the quarter, prior to the lectures on psychoanalysis. The total number tested was 119 males and 90 females. The sizes of the groups in the various testing sessions are shown in Table 1. Each session lasted approximately one hour and conditions were standard throughout. The directions given to the subjects were the same as described in the preceding chapter, with the added request that names be omitted from the test papers. For the purposes of the study, it was felt that anonymity would be preferable. The groups were not instructed to keep the nature of the experiment secret, since such prohibitions are frequently violated and thereby tend to facilitate uneven transmission of information.

TABLE 1
TEST ADMINISTRATIONS

Males		Females	
Date tested	No. in group	Date tested	No. in group
Jan. 14	30	Jan. 15	30
Jan. 20	28	Jan. 21	19
Jan. 22	27	Feb. 5	6
Jan. 27	34	Apr. 6	20
		Apr. 8	15
	119		90

The process of translating these test data into the psychoanalytic language of sex differences and dimensional intercorrelations will be the theme of the next chapter.

IV. FROM PROTOCOL TO THEORY

The gap between Blacky Test protocols and inferences concerning psychoanalytic theory is bridged by a scoring process in which an individual's responses are converted to scores along a series of psychoanalytic dimensions. The reader must continue to bear in mind that these dimensions are derived from the test and only represent analogues of the theoretical dimensions. The scoring system will be described, for purposes of exposition, in the reverse order of procedure — beginning with the final product and working backward step-by step to the original protocol.

There are 13 test dimensions for each sex. These dimensions are associated with the cartoons as follows:

Cartoon I	Oral Eroticism
" II	Oral Sadism
" III	Anal Expulsiveness
" III	Anal Retentiveness
" IV	Oedipal Intensity
" V	Masturbation Guilt
" VI	Castration Anxiety (Males)
" VI	Penis Envy (Females)
" VII	Positive Identification
" VIII	Sibling Rivalry
" IX	Guilt Feelings
" X	Positive Ego Ideal
(Fig. 11 for Males; Fig. 12 for Females)	
Cartoon XI	Narcissistic Love-Object
(Fig. 12 for Males; Fig. 11 for Females)	
Cartoon XI	Anaclitic Love-Object
(Fig. 12 for Males; Fig. 11 for Females)	

Every person ultimately receives a score on each of the 13 dimensions. The scores are recorded in terms of three categories of involvement: "Very Strong" (++); "Fairly Strong" (+); "Weak or Absent" (0). Thus, an individual's test responses are expressed as a profile of scores along the dimensions (see Table 2).

Profiles of Dimensional Scores

A subject's score (++ or + or 0) on any single dimension is based upon data from four sources: Spontaneous Story; Inquiry; Cartoon Preference; Related Comments on other cartoons. For each of these sources

TABLE 2
PROFILES OF DIMENSIONAL SCORES

Male subjects	Oral Eroticism	Oral Sadism	Anal Expuls.	Anal Retent.	Oedipal	Masturb. Guilt	Castration Anxiety	Positive Ident.	Sibling Rivalry	Guilt Feelings	Positive Ego Ideal	Narcissistic Love-Object	Anaclitic Love-Object
IM - 14	++	++	++	0	+	+	+	0	++	+	0	+	0
IM - 15	0	0	0	+	0	+	0	+	0	0	0	0	0
IM - 16	+	0	++	0	++	0	+	0	0	0	+	0	++

his responses are scored as showing either strong involvement on the dimension or not. By way of illustration, an individual's score on the dimension of Oral Eroticism is arrived at as indicated in Table 3.

TABLE 3

	Source		Oral Erotic Involvement
Subject A:	Cartoon I	Spontaneous Story	Strong
	Cartoon I	Inquiry	Not Strong
	Cartoon I	Preference	Not Strong
	Related Comments on other cartoons		Strong

These separate source scores are then combined into the final dimensional score (++ or + or 0) which appears in the profile. If a subject is scored Strong on all four or three of the four sources, he is categorized on the dimension as Very Strong (++). If he is scored Strong on two of the four sources (as in the example just cited), he is categorized as Fairly Strong (+). If he is scored Strong on one or none of the four sources, he is categorized as Weak or Absent (0).

The next procedure to consider is the classification of the sources into either Strong or Not Strong. In the Inquiry, certain responses are keyed as Strong and the number of Strong responses which a subject gives to the questions on a particular dimension is totaled. The distribution of scores for all subjects is then examined, and an arbitrary cut-off point is chosen--all scores at that point or above being classified as Strong and all those below as Not Strong. The scoring keys and cut-off points for the complete test Inquiry are presented in Appendix *B*.

The Cartoon Preference is scored on all dimensions except Positive Identification, Positive Ego Ideal, Narcissistic Love-Object, and Anaclitic Love-Object. The theoretical significance of the Preference for the latter dimensions seemed sufficiently inconclusive to warrant its omission. Where scored, the Preference is recorded as indicating Strong involvement if the

cartoon was selected by the subject as either the Best or the Worst of all the cartoons.

Related Comments on other cartoons are included in the scoring of all dimensions. While scoring the dimension of sibling rivalry, for example, the entire record, apart from Cartoon VIII, is examined for references to sibling rivalry. If *any* such references are found, the subject is given a score of Strong.

Scoring the Spontaneous Story as either Strong or Not Strong is a more complicated procedure. Fairly explicit scoring standards have been set up for the stories on each cartoon. Many sample stories from the preliminary administrations of the test are also available. However, the ability to score stories with minimum error must remain a function of thorough grounding in psychoanalytic theory plus competence to interpret projective material (such as provided by the Thematic Apperception Test), as well as close familiarity with the Blacky Test itself. In order to determine the reliability of scoring Spontaneous Stories under these prerequisite conditions, the following study was undertaken.

The author and his associate,[8] who had worked jointly in setting up the scoring standards, independently scored a series of 25 stories on each dimension. All stories were selected at random. The percent agreements between the two independent sets of scores are shown in Table 4, along with the chance expectancy. The results strongly support the hypothesis that the author's scoring of Spontaneous Stories in the experiment proper was sufficiently consistent to rule out serious concern over unreliability of scoring.

The Spontaneous Story scoring standards for each dimension are to be found in Appendix *C*. To illustrate the flavor of the stories and their scoring, verbatim samples (from different subjects) of a "Strong" and a "Not Strong" story for each dimension are given below:

Cartoon I Oral Eroticism

Strong:

> Blacky has just discovered the delightful nectar that Mama can supply—it is an endless supply and she is enjoying it. She doesn't know where it comes from, but she doesn't care. Mama is pacific throughout it all—she doesn't particularly like this business of supplying milk, but she is resigned to it. It is a pretty day and they are both calm and happy.

---

[8]The author is immeasurably indebted to H. L. Raush, who devoted countless hours to this phase of the research.

TABLE 4
SPONTANEOUS STORY SCORING AGREEMENT

Dimension	Obtained per cent scoring agreement	Per cent expected by chance
Oral Eroticism	100	56
Oral Sadism	96	61
Anal Expulsiveness	84	43*
Anal Retentiveness	96	44*
Oedipal Intensity	96	53
Masturbation Guilt	84	50
Castration Anxiety (Males)	100	56
Penis Envy (Females)	76	53
Positive Identification	100	79
Sibling Rivalry	92	56
Guilt Feelings	80	51
Positive Ego Ideal	100	92
Narcissistic Love-Object	92	54
Anaclitic Love-Object	100	85
	Obtained	Expected by chance
Mean	92.6%	59.5%
Median	96.0%	55.0%
Range	76%–100%	43%–92%

*These two dimensions were scored on a four-point scale instead of a two-point scale. Hence the chance expectancy of agreement is somewhat lower.

Not Strong:

Blacky, a male pup of a few weeks, is having his midday lunch. Mama is bored with the proceedings but as a mother with her maternal instincts is letting Blacky have his lunch to Blacky's satisfaction.

Cartoon II Oral Sadism

Strong:

Seems like Blacky has a matricidal desire. Of course, maybe he can't read, and just likes to chew and shake things. Maybe he has a fierce temper or something. Then again, maybe he's hungry, and mad because Mama isn't there to feed him.

Not Strong:

Blacky is playful, and has gotten Mama's collar away from her. She's having a wonderful time pretending it's an animate object, and she must fight with it so as to make it be quiet. The fierce look is fooling nobody but Blacky.

Cartoon III Anal Expulsiveness

Strong:

Blacky, still frustrated, shows his contempt of Mama by leaving a pile of defecation by her house. "There!" he's probably thinking, "That'll take care of her!"

Not Strong:

Blacky was not too slow when it came to housebreaking. It took him little time at all to learn that he must relieve himself either outdoors or in. Outdoors he went where the occasion demanded, unconfined and relieved.

Cartoon III Anal Retentiveness

Strong:

I suppose she is covering it up because sometime she has been scolded for it and even as a puppy feels there is something not quite public about it. That is why she chooses a spot a little away from her doghouse.

Not Strong:

All the family is out, but Blacky has gone between Mama and Papa's houses to "relieve herself," because she is still young, and feels most at ease near her parents' houses.

Cartoon IV Oedipal Intensity

Strong:

This verifies the idea that Blacky is jealous of Mama's love for Papa and the indication is that he will very soon do something to express his feeling. He will, perhaps, attempt to meet Papa as an equal and fight Papa for Mama's affection; he is rash and hard-headed and does not realize the situation is unalterable.

Not Strong:

Blacky is a little bit jealous these days. She has been in the habit of having her parents lavish all their attention on her—and then she discovered Mama and Papa behind a bush the other day and they weren't paying the slightest bit of attention to her.

Cartoon V Masturbation Guilt

Strong:

Blacky has just discovered what a nice sensation he has when he licks his sexual parts and he has no inhibitions whatever about it—licking them any time he feels the urge, yet this does not stop him from being amorous every time he sees a girl dog.

Not Strong:

Blacky is quite amazed. What has she discovered? Mama and Papa have never told her anything about this. She is quite amazed. She will have to see Mama and Papa and ask them about herself. She can't seem to comprehend this great mystery.

Cartoon VI Castration Anxiety (Males)

Strong:

Blacky doesn't like the looks of this at all, and he's very conscious of his big black tail out behind him. But his big worry is to see if this new thing that his sister is undergoing will hurt much. The expression on Tippy's face is worried, and Blacky's beginning to think that way.

Not Strong:

Tippy, Blacky's little brother, is about to have his tail chopped off unknowingly. Blacky is about to bark and warn him of his danger. Physically there is nothing he can actually do, but he will also try to push him away from the block.

Cartoon VI Penis Envy (Females)

Strong:

(Blacky's curiosity has been aroused about the opposite sex and she decides to look closely at Tippy's sex organs when he is not aware that she is looking at him.)[9] Tippy's tail is going to be cut off and Blacky watches interestedly.

Not Strong:

Blacky does not quite understand what is going to happen to Tippy but she is afraid because she senses it will not be good. She would like to help him, but doesn't know what to do.

Cartoon VII Positive Identification

Strong:

Now listen you, you little pooch, when I bark, you jump, do you get that? Blacky feels very superior to this little dog. He is making believe that he's the boss, or maybe pretending to be his father talking to him in a superior tone.

Not Strong:

Blacky has found something peculiar. It looks like a dog, but it doesn't move or bark at her. She's a little afraid of it at first, but later she realizes she's bigger and the toy can do her no harm—so she settles down to enjoy it.

Cartoon VIII Sibling Rivalry

Strong:

Blacky feels a little dejected because Tippy is getting all the attention. She is frustrated to the point of thinking of running away from home forever and ever. Her little heart is heavy and there are faint sobs coming from deep in her chest. Life is horrible, and she wishes she were never born.

Not Strong:

Blacky wonders why Tippy is getting so much attention and he is getting none. He is somewhat jealous and wondering what caused Tippy to get so much attention. He is also trying to figure out how he may get some attention.

---

[9]The parenthetical material, though crossed out by the subject, remained legible and was weighted in the scoring. This procedure was followed routinely in light of the presumed significance of crossed-out material.

Cartoon IX Guilt Feelings

Strong:

Blacky is feeling very much like killing himself. His little guardian angel has come down and is talking to him, and trying to reason with Blacky. Do we reason with our children?

Not Strong:

Blacky has just gone over to the neighbors and has brought home a pair of shoes. She did this to please, but they weren't accepted pleasantly and Blacky was scolded and the shoes taken back. This is probably the second time Blacky has done this and now knows she's naughty—hence the conscience. She doesn't like disfavor, but would rather be petted, so she won't do it again.

Cartoon X Positive Ego Ideal

Strong:

In his dream one day, Blacky dreamt of himself being a very fantastic dog—big, brave, and useful. This reminded him of his papa very much. He was very proud of it, people stared at him everywhere.

Not Strong:

He has been made fun of for being so droopy. He resents it and now pictures himself as a beautiful pointer. When he wakes up, he will really feel low and drag himself back to Mom, who will comfort him and tell him how pretty he is.

Cartoon XI—Narcissistic Love-Object

Strong:

Here Blacky visualizes himself as a lady dog—wishing that he might be changed because he feels he would make better time among the feminine circle than he can do as a male dog. Blacky probably is either awfully self-conscious of his behavior with other male dogs or feels he would like to take it easy as a lady.

Not Strong:

Wow—Blacky thinks, as he dreams of the girl he will someday woo and win. Isn't she a pip? When I meet her I'll really show her what a good man I am. Maybe we can have a family like Mama and Papa. —I want about ten instead of only two. Ah me.

Cartoon XI Anaclitic Love-Object

Strong:

Blacky is dreaming about the female of his hopes, and he is going to make love to her, and there won't be any other male dogs around to bother him, because he is going to chase them away in front of her eyes. He thinks she is just like his mother, who is the prototype of all females in his eyes.

Not Strong:

> Blacky is dreaming of his girl friend that he is going to meet when he grows up. He will probably meet her in the local bar, dance a few times with her and then she will be madly in love with him from then on till now.

The reader may have observed, both from Appendix *C* and the samples, that the scoring of a story is not based primarily upon the manifest content. The scoring standards have been attuned to pick up the subtle hints from which pertinent clinical inferences can be drawn. For example, if a person completely ignores the anal reference on Cartoon III ("Here Blacky is relieving himself") and proceeds to write a story about Blacky burying a bone, this is construed as a case of repression, and consequently he is given a score of Strong. If he expresses a denial (see sample of Strong story on Cartoon V), again he is scored Strong. Likewise he is penalized if he blocks or is extremely evasive, or obviously produces an artificially happy story, makes significant slips, or indulges excessively in various other defense mechanisms such as rationalization, projection, displacement, and so on. In short, the scoring of the Spontaneous Story is directed, not toward what is actually written down, but rather toward the underlying implications or latent content. Attempts by a subject to mollify his stories seem readily detectable by this approach.

The preceding sections have described how the four sources of data are each classified into Strong or Not Strong. The mode of combining them into dimensional scores (++ or + or 0), already illustrated in the case of Oral Eroticism, is presented in Appendix *D* for all dimensions in detail. The next stage in relating responses in the protocol to psychoanalytic theory involves statistical analyses of the data thus obtained.

According to the design of the experiment, consonance between test findings and psychoanalytic theory is to be examined in two general areas: (*a*) sex differences; and (*b*) dimensional intercorrelations. In the area of sex differences, the male and female subjects were compared, by the chi-square technique, in regard to total score on each dimension; answers to items in the Inquiry; and the number ignoring the introductory comments on Cartoons III (Anal Sadism) and V (Masturbation Guilt). The level of statistical significance was arbitrarily set at .05 or less (5 per cent level). Twenty-three of the chi-square comparisons met this criterion. Table 5 shows that the obtained number of statistically significant chi-squares far exceeds the chance expectancy.

In the second area of the study, interrelationships of the test dimensions,

TABLE 5
NUMBER OF SIGNIFICANT SEX DIFFERENCES OBTAINED VS. EXPECTED
(Total number χ^2 possible = 121)

Level of significance	Number obtained	Number expected by chance alone
5% or less	23.0	6.1
2% or less	21.0	2.4
1% or less	15.0	1.2

a complete table of intercorrelations was prepared. Total scores on each dimension were correlated with total scores on every other dimension by using the tetrachoric method, with the data dichotomized into Very Strong (++) and Fairly Strong (+) as one category versus Weak or Absent (0) as the other category. The intercorrelations were computed for males and females separately as well as combined. The level of statistical significance of each relationship was ascertained by the chi-square technique, with $P < .05$ as the criterion of significance. Examination of Table 6 reveals again that the number of statistically significant findings far exceeds the chance expectancy.

TABLE 6
NUMBER OF SIGNIFICANT DIMENSIONAL INTERCORRELATIONS OBTAINED VS. EXPECTED
(Total number of r_t's possible = 225)

Level of significance	Number obtained	Number expected by chance alone
5% or less	28.0	11.3
2% or less	18.0	4.5
1% or less	13.0	2.3

Thus, with Tables 5 and 6, the process of linking Blacky Test findings and psychoanalytic theory begins to unfold. Many statistically significant test results have been obtained. These results cannot be accounted for in terms of chance expectancy alone. While any single test finding may, of course, be a chance occurrence, the total setting demands further investigation. Can the results be explained away as test artifacts? If not, do they agree or disagree with the theory? Answers to these questions will be sought in the chapters which follow.

V. DIFFERENCES BETWEEN THE SEXES

Blacky Test data have revealed numerous areas of sex differences which satisfy a criterion of statistical significance.[9a] Explanation in terms of the operation of chance factors alone has been ruled out. The next step is to examine each of the specific differences in regard to its consistency with the psychoanalytic theory of psychosexual development. In accordance with the design of the experiment (Section III), only those test results which possess statistical significance are to be evaluated in terms of agreement or disagreement with the theory. All of these obtained differences will be investigated also for possible explanations other than those suggested by the theory. Each area of differences between the sexes, found from the Blacky Test, will be discussed in the following outline form:

Area of sex differences on the Blacky Test.

1. Evidence from psychoanalytic theory.[10]
2. Evidence from Blacky Test.
3. Interpretation of 1 and 2.

The evidence from psychoanalytic theory (Section 1) will be direct quotations, wherever possible, from Sigmund Freud's own writings and from Otto Fenichel's (1) textbook on psychoanalytic theory which appeared in 1945. In some cases direct evidence will be lacking; for these the author will seek to provide indirect or inferential material. In a few there may be no theoretical sources available at all. The Blacky Test sources (Section 2) will be presented explicitly, along with the level of statistical significance. The interpretation of evidence (Section 3) will be an attempt to collate and evaluate the separate sources, including possible extraneous influences such as test artifacts.

A. Prevalence of Oral Sadism

1. *Psychoanalytic Theory*

Fenichel maintains that the little girl, as a result of pre-oedipal frustrations by the mother, feels hostile toward her and turns away from the mother to the father. He says:

> The most important experiences that precipitate, facilitate, impede, or form the change of object in girls are disappointments coming from the

[9a]Those areas in which no significant sex differences were found are given in Appendix *E*.

[10]For purposes of exposition, the theoretical evidence is presented before the Blacky Test data. The actual procedure was the reverse.

> mother, which cause a turning away from her. Among these, weaning, training for cleanliness, and the birth of siblings have the most important repercussions. . . .(1, p. 89).

The specifically feminine disappointment accounting for the greater antagonism in females is the lack of a penis, for which the mother is held responsible. Awareness of this lack remobilizes anal and especially oral elements, which are often characteristic of subsequent femininity. Fenichel adds:

> It is understandable that this development is open to many disturbances and that conflicts about the pre-oedipal love for the mother play an important rôle in the neuroses of women. In normal development, too, the relationship of women to their mothers is more frequently ambivalent than is that of most men to their fathers. Some remnants of the pre-oedipal mother fixation are always found in women. . . .(1, p. 90).

Freud links the little girl's hostility directly to oral frustrations:

> The [girl's] complaint against the mother that harks back furthest, is that she has given the child too little milk, which is taken as indicating a lack of love. . . .(6, p. 166).

Thus, psychoanalytic theory would predict oral sadism, which is an expression of pre-oedipal hostility toward the mother, to be more prevalent in females than in males. An interesting corollary is the relationship advanced between oral sadism and manic-depressive psychoses. Fenichel reproduces a chart drawn up by Abraham, in which manic-depressive disorders are described as having their dominant point of fixation in the oral-sadistic stage. Manic-depressive disorders are found to be more common among women.

2. *Blacky Test*

Significantly more females than males (2 per cent level) chose the oral-sadistic alternative (3) in answer to the following question on Cartoon II:

> What will Blacky do next with Mama's collar?
> (1) Get tired of it and leave it on the ground.
> (2) Return it to Mama.
> (3) Angrily chew it to shreds.

3. *Interpretation*

With respect to the prevalence of oral sadism, the Blacky Test finding supports the theoretical expectation that more females than males retain oral-sadistic tendencies.

B. Oral "Voracity"

1. *Psychoanalytic theory*

The theoretical sources do not offer any direct or indirect evidence concerning a possible sex difference in greediness.

2. *Blacky Test*

Significantly more males than females (5 per cent level) picked the voracious alternative (1) in answer to the following question on Cartoon I:

> Which one of the following best describes Blacky?
> (1) He's (she's) a little glutton who never stops eating.
> (2) He's (she's) got a hearty appetite which usually gets satisfied.
> (3) He (she) sometimes doesn't get enough to replace all the energy he (she) burns up.

3. *Interpretation*

The absence of a theoretical viewpoint in this connection precludes evaluation. Nor can the test finding be interpreted with any degree of confidence, since females may have merely refrained from choosing the alternative more often because of its "un-ladylike" quality.

C. Repression of Anal Sadism

1. *Psychoanalytic theory*

There is no explicit theoretical evidence concerning a sex difference in repression of anal sadism. However, a difference can be inferred from Freud's analysis of "hostile and obscene wit." He equates obscene jokes with anality in the following passage:

> The sexual element which is at the basis of the obscene joke comprises more than that which is peculiar to both sexes, and goes beyond that which is common to both sexes, it is connected with all these things that cause shame, and includes the whole domain of the excrementitious. However, this was the sexual domain of childhood, where the imagination fancied a cloaca, so to speak, within which the sexual elements were either badly or not at all differentiated from the excrementitious. In the whole mental domain of the psychology of the neuroses, the sexual still includes the excrementitious, and it is understood in the old, infantile sense (8, pp. 693-4).

Then he proceeds to comment on the function of obscene humor and its differential effect upon the sexes:

> It now becomes comprehensible what wit accomplishes through this service of its tendency. It makes possible the gratification of a craving (lewd or hostile) despite a hindrance which stands in the way; it eludes the hindrance and so derives pleasure from a source that has been inaccessible on account of the hindrance. The hindrance in the way is really nothing more than the higher degree of culture and education which correspondingly increases the inability of the woman to tolerate stark sex matters. . . . The power which makes it difficult or impossible for the woman, and in a lesser degree for the man, to enjoy unveiled obscenities we call "repression" . . .(8, p. 696).

In another reference he alludes to the greater repression of aggressiveness in women:

> The repression of their aggressiveness, which is imposed upon women by their constitution and by society. . . .(6, p. 158).

These statements suggest, in a rather roundabout fashion, that more females than males might be expected to repress anal-sadistic tendencies.

2. *Blacky Test*

a. Significantly more females than males (1 per cent level) remained oblivious to the introductory comment on Cartoon III ("Here Blacky is relieving himself [herself]") and gave non-anal stories. In terms of the test, this avoidance is construed as repression.

b. Significantly more males than females (1 per cent level) selected the anal-sadistic alternative (4) in answer to the following question on Cartoon III:

> How does Blacky feel about the training he's (she's) been getting?
> (1) By relieving himself (herself) in the way he's (she's) been taught, he (she) now has an opportunity to show his (her) family what a good dog he (she) can be.
> (2) He (she) feels Mama and Papa are expecting too much of him (her) at this early stage.
> (3) He (she) is very happy to have control of himself (herself).
> (4) He (she) thinks he's (she's) got Mama and Papa right where he (she) wants them.

3. *Interpretation*

The test findings, like the theory, suggest more extensive repression among females. This is particularly true of the evidence from the

Spontaneous Story. The item from the Inquiry, while conceivably indicating greater repression in the case of the females, might possibly be a reflection of originally more widespread anal sadism among males. The latter source therefore cannot be given as much weight as the Spontaneous Story. Evidences from the culture, whose influence Freud specifically mentions in this connection, tend to reinforce the notion that males are freer to express aggressions in an anal context, e.g., swearing by using anal terms. The intrepid reader might perform his own validation of this point by observing differences between the sexes in reactions to anal-sadistic references.

D. Retention of Pre-Oedipal Types of Object Relationships

1. *Psychoanalytic Theory*

In the course of psychosexual development, there is normally an advance from the early pre-oedipal type of object relationship, in which the child passively experiences a narcissistic need to be loved, to the more active love-seeking which begins in the oedipal period. Typically, residues of the more primitive fear of losing love are retained in varying extents during the later stages. Psychoanalytic theory holds that these residues are more prominent in the case of the female, as a result of the change of object in girls. When the boy reaches the oedipal phase, he continues his original relationship with the mother; the girl must renounce her maternal ties in favor of the father. The completeness of the change is therefore less likely in females than in males. Excerpts from Fenichel on this topic are:

> The fact that in this type of identification there is also a regressive element in love is clearer in women than in men. The passive aim of female sexuality is more closely related to the original aims of incorporation than is the active aim of male sexuality. Therefore, passive sexuality has more archaic features than active sexuality. The aim of being loved is more stressed in women than the aim of loving—the narcissistic need and the dependency on the object are greater (1, p. 85).
>
> . . . Some remnants of the pre-oedipal mother fixation are always found in women. There are many women whose masculine love objects have more characteristics of their mother than of their father (1, p. 90).
>
> . . . analysis shows that other and older fears, above all the fear over loss of love, are stronger in women. . . .(1, p. 99).

Freud stresses the retention of pre-oedipal attachments in females in the following comments:

> . . . The girl's Oedipus complex has long concealed from us the pre-

> oedipal attachment to her mother which is so important and which leaves behind it such lasting fixations. . . . (6, p. 176).
>
> . . . Regression [in women] to fixations at these pre-oedipal phases occur very often. . . . (6, p. 179).
>
> . . . In short, we gain the conviction that one cannot understand women, unless one estimates this *pre-oedipal attachment to the mother* at its proper value (6, p. 163).

2. *Blacky Test*

Evidence pertaining to the retention of pre-oedipal types of object relationships is derived from Cartoon IV, in which both pre-oedipal and oedipal responses are possible.

a. More males than females (1 per cent level) select the oedipal alternative (1), more and more females than males (1 per cent level) select the pre-oedipal alternative (2) in answer to the following question:

> Which one of the following makes Blacky most unhappy?
> (1) Papa keeping Mama all to himself (Mama keeping Papa all to herself).
> (2) The idea that Mama and Papa seem to be ignoring him (her) on purpose,
> (3) He (she) is ashamed watching them make love out in the open.

b. Significantly more females than males (1 per cent level) prefer to be with the same-sex parent:

> Which would make a happier picture?
> (1) Mama left on the outside watching Blacky together with Papa.
> (2) Papa left on the outside watching Blacky together with Mama.

c. Significantly more males than females (5 per cent level) obtain strong total scores on the dimension of Oedipal Intensity.

3. *Interpretation*

Theory and test are in specific agreement concerning the greater retention of pre-oedipal components in female than male object relationships.

E. Sadistic Conception of the Sex Act

1. *Psychoanalytic Theory*

Freud discusses this area as follows:

> The third of the typical sexual theories appears in children when through some unforeseen domestic occurrence they witness parental sexual intercourse, concerning which they are then able to obtain only a very incomplete idea. Whatever detail it may be that comes under their observation, whether it is the position of the two people, or the sounds, or certain accessory circumstances, in all cases they arrive at the same conclusion, that is, at what we may call the *sadistic conception of coitus*, seeing in it something that the stronger person inflicts on the weaker by force, and comparing it, especially the boy, to a fight as they know it from their childish play. . . . (3, p. 69).

Unfortunately the above statement is ambiguous in its implications for a study of sex difference. It is not clear whether Freud means that the boy is only more likely to compare parental intercourse to a fight or whether he is also more likely to conceive of it as sadistic.

2. *Blacky Test*

Significantly more males than females (2 per cent level) chose the sadistic alternative (3) in answer to the following question on Cartoon IV:

> What does Blacky suspect is the reason behind the scene he's (she's) watching?
>
> (1) He (she) suspects Mama and Papa are planning an addition to the family.
> (2) He (she) suspects Mama and Papa are very much in love.
> (3) He (she) suspects Papa is having his own way about things.
> (4) He (she) suspects Mama and Papa are purposely depriving him (her) of attention.

3. *Interpretation*

Agreement between psychoanalytic theory and Blacky Test cannot be evaluated satisfactorily on this point, inasmuch as the available source material for the theoretical position lacks any clear definition. The test finding suggests that males may be more likely than females to regard sexual intercourse as a sadistic act. It is also possible that the answers to the questions are merely an expression of masculine feelings of dominance rather than of a specific conception of the sex act.

F. Prevalence of Masturbation Guilt

1. *Psychoanalytic Theory*

No evidence has been found from the writings of either Freud or Fenichel which bears on the question of whether masturbation guilt is more prevalent among males or females.

2. *Blacky Test*

a. Significantly more males than females (1 per cent level) selected the guiltless alternative (1) in answer to the following question on Cartoon V:

How does Blacky feel here?
(1) Happy, without a care in the world.
(2) Enjoying himself (herself), but a little worried.
(3) Mixed up and guilty.

b. Significantly more males than females (2 per cent level) also picked the guiltless alternative (1) in answer to the following question on Cartoon V:

How might Blacky feel about this situation when he (she) is older?
(1) Happy, without a care in the world.
(2) Enjoying himself (herself), but a little worried.
(3) Mixed up and guilty.
(4) The situation won't come up again when he (she) is older.

3. *Interpretation*

This area remains in the realm of speculation. The theory furnishes no information, and the test findings are capable of two opposing interpretations. On the one hand, the results may be taken literally to mean that more females than males possess masturbation guilt. Support for this view stems from the cultural fact that the topic of masturbation tends to be more secretly guarded from discussion by females than by males, and hence possibly is more likely to generate guilt. On the other hand, since these two items are the least disguised, the males conceivably may be selecting the guiltless alternatives in an attempt to deny guilt, which may in actuality be more prevalent among men. The theoretical linking of masturbation guilt with fears of castration and oedipal fantasies may be considered consonant with this position.

G. Identification with Parent of the same Sex

1. *Psychoanalytic Theory*

In terms of psychoanalytic theory, the resolution of the Oedipus complex is normally accompanied by an identification process, in which the boy represses his sexual desires for the mother and patterns himself after his powerful rival, the father, whereas the girl gives up her strivings toward the father and patterns herself after the mother. The normal identification therefore is with the parent of the same sex. The suddenness and completeness with which this change takes place differs for the two sexes. Fenichel's discussion of this point follows:

> . . . The boy gives up his sensual and hostile Oedipus wishes because of a castration fear, the intensity of which is due to the hypercathexis of the penis during the phallic phase. The complex, according to Freud, "is smashed to pieces by the shock of threatened castration." In girls, however, it is given up because of fear over loss of love, because of disappointment, shame, and also fear over physical injury. All of these forces are of a lesser dynamic value than castration fear; thus the passing of the Oedipus complex in girls generally comes about in a more gradual and less complete way. . . . (1, p. 108).

2. *Blacky Test*

Evidence concerning parental identifications is obtained from answers to the following four questions on Cartoon VII:

a. "Who talks like that to Blacky — Mama or Papa or Tippy?"— The males tend to say "Papa" and the females "Mama" (1 per cent level).

b. "Whom is Blacky imitating here — Mama or Papa or Tippy?"— The males tend to pick "Papa" and the females pick "Mama" (1 per cent level).

c. "Whom would Blacky rather pattern himself (herself) after — Mama or Papa or Tippy?" — The males say "Papa," the females "Mama" (1 per cent level).

d. "Blacky's disposition, actually, is most like the disposition of which one — Mama or Papa or Tippy?" — The males pick "Papa," the females pick "Mama" (1 per cent level).

3. *Interpretation*

Psychoanalytic theory and Blacky Test results agree that identification normally occurs with the parent of the same sex.

H. Confusion in Identification Process

1. *Psychoanalytic Theory*

Previously quoted sources (Areas $A_1D_1G_1$) have pointed to differences between the sexes in regard to completeness of the identification process. Females are said to remain more ambivalent toward their mothers than males do toward their fathers; they retain more pre-oedipal elements, which interfere with subsequent development; the change of objects in girls impedes the completeness of the oedipal phase; and the resolution of the Oedipus complex is more gradual and less complete in the case of females. For these reasons, the process by which the female identifies with her mother is considered to be less clear-cut than the process whereby the male identifies with his father. In this connection Freud says:

> The mother-identification of the woman can be seen to have two levels, the pre-oedipal, which is based on the tender attachment to the mother and which takes her as a model, and the later one derived from the Oedipus-complex, which tries to get rid of the mother and replace her in her relationship with the father. Much of both remains over for the future. . . . (6, p. 183).

2. *Blacky Test*

Two items from Cartoon VII bear on this topic:

a. "Whom would Blacky rather pattern himself (herself) after — Mama or Papa or Tippy?"—Although the males tend to say "Papa" and the females "Mama," the males are significantly more *decisive* than the females in making their choices; i.e., a greater percentage of males choose "Papa" than females choose "Mama" (1 per cent level).

b. "Whom is Blacky most likely to obey — Mama or Papa or Tippy?"—Both males and females say "Papa" significantly more than "Mama" (1 per cent level).

3. *Interpretation*

The test findings support the theoretical view that the identification process is less clear-cut in the case of females. The latter are less sure with which parent they seek to identify. The process is further complicated by the fact that the father is regarded by females as a figure more likely to be obeyed than the mother — the theory being that the decisive frustrating agent is a crucial influence in identification (1).

I. Aggression Harbored toward Identified Parent

1. *Psychoanalytic Theory*

The theoretical position has already been discussed in the preceding section (Area H). Females are said to retain stronger undercurrents of hostility toward their identified parent, the mother, than males do toward the father. Freud makes the following statement:

> The turning away from the mother in females occurs in an atmosphere of antagonism; the attachment to the mother ends in hate. Such a hatred may be very marked and may persist throughout an entire lifetime; it may later on be carefully overcompensated; as a rule, one part of it is overcome while another persists. . . . (6, p. 165).

2. *Blacky Test*

Significantly more females than males (2 per cent level) selected the aggressive, hostile alternative (4) in answer to the following question on Cartoon VII:

> What would Blacky have an impulse to do if he (she) were in the position of the toy dog?
>
> (1) Get frightened and hide.
> (2) Stand there and take it.
> (3) Get mad and sulk.
> (4) Start fighting.

3. *Interpretation*

On Cartoon VII, which is intended to portray the identification process, the subject (Blacky) presumably behaves toward the toy dog in the same manner as the identified parent behaved toward him (her) in real life. The above question from the Inquiry reverses these rôles, with the subject now in his original subservient position. Reactions to this situation are taken to indicate the underlying attitude of the subject toward the parent with whom he has sought to identify himself. Consequently, the test finding is consistent with the theory, since more females do express hostile impulses toward the identified parent.

J. Prevalence of Guilt Feeling

1. *Psychoanalytic Theory*

Neither Freud nor Fenichel seems to offer a statement concerning the relative prevalence of guilt feelings in males and females. A sex difference may be deduced from Fenichel's description of the genesis of guilt, but it

must be recognized that such an extension is a tenuous one. Fenichel links guilt feelings to fear of losing love or narcissistic supplies in the following manner:

> When the ego is sufficiently developed to form a judgment that there is danger of a cessation of essential narcissistic supplies, the aim of its signal, "annihilation may occur," must be to influence objects to furnish these supplies. This state represents the anxiety over loss of love which plays so important a rôle as a motive for defense. . . .
>
> . . . The anxiety over loss of love, or rather the anxiety arising out of loss of narcissistic supplies, turns into anxiety over loss of the superego's supplies, and the fear into guilt feeling. . . . (1, pp. 135-6).
>
> Guilt feeling not only has an oral character in general but an oral-sadistic character in particular. . . . (1, p. 138).

Thus, guilt arises from fear of losing the love of the superego introjects or "internalized" parental providers of narcissistic supplies. Concern over loss of love is said to be more prevalent in woman than in men. The latter are more preoccupied by external fears, which have their prototype in the fear of castration. Physiological and anatomical differences between the sexes also play some part in the divergent development. On this point Fenichel says:

> No doubt this physiological difference [in performance of sexual intercourse] contributes to the prevalent rôles of castration fear or fear over loss of love in man and woman respectively. However, this cannot be more than a relatively late secondary contribution. The relative preponderance of the respective fears is established in childhood, long before the first experiences in sexual intercourse (1, p. 100).

From these observations it may be deduced that females, who have more conflicts over internal deprivations and loss of love, theoretically should show greater evidence of guilt feelings than males do.

2. *Blacky Test*

a. Significantly more females than males (2 per cent level) obtained strong total scores on the dimension of guilt feelings.

b. Significantly more males than females (1 per cent level) chose the externalized alternative (3) in answer to the following question on Cartoon IX:

> How is Blacky's conscience here?
>
> (1) His (her) conscience is so strong he's (she's) practically paralyzed.

(2) His (her) conscience is bothering him (her) somewhat, but he's (she's) mostly afraid of what will be done to him (her).
(3) He's (she's) hardly bothered at all by his (her) conscience, just afraid of what will be done to him (her).

c. Significantly more females than males (1 per cent level) expressed lasting guilt (1) in answer to the following question on Cartoon IX:

Do you think Blacky will
(1) Have this feeling as long as he (she) lives?
(2) Feel badly every now and then?
(3) Feel badly for a little while and then go out to play?

3. *Interpretation*

If the theoretical extension made in Section 1 is a legitimate one, the test findings can be said to support it. The responses indicate that females probably possess stronger guilt, whereas males are more concerned with fears of external harm.

K. Figures Introjected as Superego

1. *Psychoanalytic Theory*

Psychoanalytic theory holds that the superego attains its major development, with the passing of the Oedipus complex, in a process whereby parental standards and prohibitions are incorporated or "introjected" by the child as his own. Fenichel's discussion of this point follows:

> . . . It is true that in accordance with the "completeness" of the Oedipus complex, everyone bears features of both parents in his superego. Under our cultural conditions, however, generally for both sexes the fatherly superego is decisive; in women, moreover, a motherly superego is effective as a positive ego ideal. . . . (1, p. 104).
>
> . . . However, the sexual differences in the formation of the superego are certainly not the same under different cultural circumstances. . . . (1, p. 469).

Hence, the superegos of both males and females are expected to show more characteristics of the father than of the mother, though cultural circumstances may alter this pattern.

2. *Blacky Test*

Although the proportions are mixed for both sexes, significantly more males than females attribute fatherly characteristics to the superego, where-

as significantly more females than males attribute motherly characteristics to the superego (1 per cent level) in answer to the following question on Cartoon IX: "Which character do the actions of the pointing figure remind Blacky of?"

3. *Interpretation*

Fenichel's analysis of the culture as conducive to the formation of fatherly superegos in women as well as in men is not confirmed by the test data. From the Blacky Test, women tend to have motherly superegos. The cultural implications in this area are necessarily limited by the fact that the female subjects in the experiment cannot be considered representative of American women. Recent emphasis on the growing influence of the mother in family life is consistent with the test finding for the females, however. The theoretical observation that the superego in both sexes contains mixed parental elements is supported by the fact that 31 per cent of the males do attribute motherly characteristics to it, while 29 per cent of the females say the superego figure in the cartoon is reminiscent of "Papa."

L. Reactions to Frustration

1. *Psychoanalytic Theory*

No references to sex differences in type of reaction to frustration have been found in the theoretical sources.

2. *Blacky Test*

Significantly more females than males (2 per cent level) select the extrapunitive alternative (2), whereas significantly more males than females (1 per cent level) select the impunitive alternative (3) in answer to the following question on Cartoon IX:

Who is really to blame for Blacky's feeling this way?
(1) Himself (herself).
(2) Somebody else.
(3) The situation couldn't be helped.

3. *Interpretation*

Agreement between theory and test cannot be evaluated in this area. The test results per se suggest that the females in the experimental sample, when frustrated, are more likely to project blame on to others (extrapunitive), while the males show a greater tendency to gloss over the situation (impunitive).

M. Attitudes Toward Attaining Ego Ideal

1. *Psychoanalytic Theory*

Again there seems to be no theoretical evidence relevant to differences between the sexes in attitudes toward attaining the ego ideal, i.e., whether males or females are more hopeful of measuring up to the positive standards which they have adopted as goals.

2. *Blacky Test*

Significantly more females than males (1 per cent level) chose the pessimistic alternative (3) in answer to the following question on Cartoon X (Figure 11 for males; Figure 12 for females):

> Actually, what are Blacky's chances of growing up to be like the figure in his (her) dream?
> (1) Very good.
> (2) Fair.
> (3) Very poor.

3. *Interpretation*

The absence of theoretical evidence precludes evaluation in this area also. Interpretation of the test finding, which reveals females to be less hopeful than males of reaching the stature of their ego ideal, must be qualified. At first glance, it seems that the females in this study are more willing than the males to adopt a realistic attitude toward their shortcomings. However, the difference may possibly be attributable to dissimilar conceptions of the ego ideal by the two sexes. Perusal of the female responses in the Spontaneous Story suggests that they are largely concerned with the physical beauty of the ideal cartoon figure, and it may be that the sex difference reflects the degree to which females have reconciled themselves to their own physical charms.

N. Prevalence of Narcissism in Love-Object Relationships

1. *Psychoanalytic Theory*

Psychoanalytic theory specifically states that narcissism in love-object relationships is more prevalent among females than males. Freud, in this connection, says:

> Further, the comparison of man and woman shows that there are fundamental differences between the two in respect to the type of object-choice, although these differences are of course not universal. Complete object-love of the anaclitic type is, properly speaking, characteristic of

> the man. . . . A different course is followed in the type most frequently met with in women, which is probably the purest and truest feminine type. With the development of puberty the maturing of the female sex organs, which up till then have been in a condition of latency, seems to bring about an intensification of the original narcissism, and this is unfavorable to the development of a true object-love. . . . (5, p. 45-6).
>
> . . . We attribute to women a greater amount of narcissism (and this influences their object-choice) so that for them to be loved is a stronger need than to love. . . . (6, p. 180).
>
> The conditions of object-choice in women are often enough made unrecognizable by social considerations. Where that choice is allowed to manifest itself freely, it often occurs according to the narcissistic ideal of the man whom the girl would have liked to be. . . . (6, p. 181).

Fenichel likewise says:

> . . . The aim of being loved is more stressed in women than the aim of loving—the narcissistic need and the dependency on the object are greater (1, p. 85).

2. *Blacky Test*

Significantly more females than males (1 per cent level) obtained strong total scores on the dimension of Narcissistic Love-Object.

3. *Interpretation*

The Blacky Test finding supports the theoretical contention that narcissistic love-object relationships are more characteristic of females than of males.

O. Summary

A summary of the data included in this chapter is now in order. The Blacky Test has revealed 14 areas in which the responses of males and females differ significantly. Evidence from the theoretical sources has been available *specifically* in seven of these areas (*A, D, G, H, I, K, N*); *inferentially* in two (*C, J*); and *not at all* in five (*B, E, F, L, M*). One conclusion immediately offers itself: Psychoanalytic theory, as presented by Freud and Fenichel, is incomplete in its treatment of sex differences in psychosexual development. Freud himself recognized this lack in a paper on "The Passing of the Oedipus Complex," written in 1924:

> It must be confessed, however, that on the whole our insight into these processes of development in the girl is unsatisfying, shadowy, and incomplete (3, p. 275).

But what about those areas in which the theory *does* take a stand? The Blacky Test data support psychoanalytic theory in six of the seven areas in which there is specific theoretical evidence, and in both of the areas in which the evidence is inferential. The sole area of disagreement is the test finding that females tend to have motherly rather than fatherly superego. This departure from Fenichel's stated opinion may very possibly be a reflection of the increasing influence of the mother in American life, in contrast to the patriarchal European society in which psychoanalysis grew up. On the whole, the consonance of theory and test in the areas of sex difference is striking. It remains to be seen whether the syntax of the theory, i.e., the relationships postulated between the dimensions, will receive equal support from the Blacky Test findings.

VI. THE SYNTAX OF PSYCHOANALYTIC THEORY

The Blacky Test has also provided a series of statistically significant tetrachoric correlations between the various test dimensions.[11] As in the case of sex differences, these syntactical relationships cannot be attributed to the operation of chance factors alone (see Table 6). Are they, then, consistent with the syntax of psychoanalytic theory? Each of the significant intercorrelations will be examined in outline form similar to that used in the preceding chapter:

Significant dimensional intercorrelation on the Blacky Test.
1. Evidence from psychoanalytic theory.
2. Evidence from Blacky Test.
3. Interpretation of 1 and 2.

The theoretical sources again will be the writings of Freud and Fenichel, quoted directly wherever possible. The Blacky Test finding for each relationship will be presented both as a *tetrachoric correlation* (r_t) and as a *chi-square level of significance* (P). The complete tables of intercorrelations are given for males alone in Appendix F; for females alone in Appendix G; and for combined males and females in Appendix H. As before, the interpretation sections will attempt to collate the evidence from theory and test, including other possible explanations of the test findings.

A. General Interrelationship of the Psychosexual Stages

1. *Psychoanalytic Theory*

According to psychoanalytic theory, there is an essential relationship among the various stages of development, e.g., oral, anal, oedipal. A gradual transition occurs from one phase to the next, with considerable overlap. Remnants of earlier levels accompany later advances. Disturbances at one stage impede the successful completion of subsequent stages, and later disturbances facilitate a return to earlier developmental levels. Freud, for example, stresses the interrelationship of anality and orality in the following sentence: "The sadistic-anal organization can easily be regarded as a continuation and development of the oral one . . . " (4, p. 590).

Excerpts from Fenichel elaborate on the topic:

> . . . all phases gradually pass into one another and overlap (1, p. 62).
>
> . . . It was stated that in mental development the progress to a higher level never takes place completely; instead characteristics of the earlier

[11]The statistically insignificant correlations are available in Appendices F, G, and H.

> level persist alongside of or behind the new level to some extent. Disturbances of development may occur not only in the form of a total arresting of development but also in the form of retaining more characteristics of earlier stages than is normal. When a new development meets with difficulties, there may be backward movements in which the development recedes to earlier stages that were more successfully experienced. Fixation and regression are complementary to each other. Freud used the simile of an advancing army in enemy territory leaving occupation troops at all important points. The stronger the occupation troops left behind, the weaker is the army that marches on. If the latter meets a too powerful enemy force, it may retreat to those points where it had previously left the strongest occupation troops. The stronger a fixation, the more easilly will a regression take place if difficulties arise (1, p. 65).
>
> Whatever the physiology of erogeneity may be, from a psychological point of view it must be said: there is not a specific oral libido, anal libido, and genital libido; there is but one libido which may be displaced from one erogeneous zone to another. . . . (1, p. 74).
>
> . . . Since parents who have had difficulty in helping their children to make an adequate adjustment on an oral level usually have difficulties again in training their children for cleanliness, and since, among the anal frustrations, the prohibition of taking feces into the mouth is especially stressed, it is understandable that anal conflicts, too, may be expressed by children through oral inhibitions, through inhibitions in eating, as well as through inhibitions in speech. . . . (1, p. 176).

Hence, persons who experience difficulties at one level are more likely to show evidences of disturbed development at others.

2. *Blacky Test*

a. Oral Eroticism vs. Oral Sadism.

(Males: $r_t = .44$)
(Females: $r_t = .30$)
Combined: $r_t = .37$, $P < .01$

Those subjects who are more disturbed than others at the oral-erotic level, i.e., receive strong scores on the dimension of Oral Eroticism, tend to be among the more disturbed at the oral-sadistic level.

b. Oral Eroticism vs. Anal Expulsiveness.

(Males: $r_t = .21$)
(Females: $r_t = .31$)
Combined: $r_t = .26$, $P < .05$

Those who show greater evidence of difficulty in the oral-erotic period also tend to score more strongly than others on the anal-expulsive dimension.

c. Oral Eroticism vs. Oedipal Intensity.

(Males: $r_t = .29$)
(Females: $r_t = .26$)
Combined: $r_t = .28$, $P < .02$

Those who possess a stronger oral-erotic component again tend to be more disturbed than others in the oedipal period.

d. Oral Sadism vs. Oedipal Intensity.

Males: $r_t = .36$, $P < .02$

Those males who receive strong scores on the dimension of Oral Sadism tend to receive strong scores on Oedipal Intensity.

e. Anal Expulsiveness vs. Oedipal Intensity.

Females: $r_t = .38$, $P < .05$

Those females who evidence greater difficulty than others at the anal-expulsive level tend to score more strongly on Oedipal Intensity.

3. *Interpretation*

The test findings, in general, support the theoretical contention that the difficulties of the various stages of psychosexual development are positively correlated with one another. The scoring on each dimension was completely independent (different sources) of the scoring on every other dimension, so the obtained correlations are not due to artifacts within the scoring system. In accordance with the design of the experiment, only those correlations which reach statistical significance are considered. Where similar trends in both males and females reach significance in combination, the data for each sex are given parenthetically.

B. Oral Eroticism vs. Castration Anxiety (Males)

1. *Psychoanalytic Theory*

The theory maintains that there is an intimate connection between castration anxiety and earlier anxieties characteristic of disturbed oral development. In Fenichel's words:

> . . . its [castration anxiety] forerunners in oral and anal anxieties over loss of breast or feces. . . . (1, p. 77).

> . . . Persons with oral fixations may fear that the penis will be bitten off, which results in confused ideas made up of both oral and genital elements (1, p. 78).
>
> It has been mentioned that anxiety over being eaten or over being bitten may be a disguise for castration anxiety. In such a case the castration fear has become distorted in a regressive way, that is, by choosing as a substitute an archaic autonomous fear. The regression may be a partial one, and frequently we see manifestations of anxiety that contain elements of both being eaten and being castrated. . . . (1, 199-200).

2. *Blacky Test*

Oral Eroticism vs. Castration Anxiety

$$\text{Males: } r_t = .35,\ P < .05$$

Those who show disturbances of the oral-erotic component are more likely to possess castration fears.

3. *Interpretation*

Theory and test are in specific agreement concerning the positive correlation between orality and castration anxiety in males.

C. Oral Eroticism vs. Positive Ego Ideal (Males)

1. *Psychoanalytic Theory*

Neither Freud nor Fenichel discusses the influence of oral fixations upon the formation of a positive ego ideal in males. However, an inferential relationship can be drawn without much difficulty. A positive ego ideal implies a boy's internalized conception of a benevolent, protecting figure, usually modeled after the father. This introjected ideal figure sets positive standards and goals which spur the youth to constructive effort. Strong oral conflicts, on the other hand, imply passive-dependent preoccupation over getting narcissistic supplies, typically from a frustrating mother. Therefore, a negative correlation between oral eroticism and positive ego ideal might be inferred.

2. *Blacky Test*

a. Oral Eroticism vs. Positive Ego Ideal.

$$\text{Males: } r_t = -.40,\ P < .01$$

Those males who are more involved with oral conflicts are less likely to have positive ego ideals.

3. *Interpretation*

The Blacky Test finding is consistent with the inference from the theory that there is an inverse relationship between orality and the formation of a positive ego ideal.

D. Oral Sadism vs. Anal Retentiveness (Males)

1. *Psychoanalytic Theory*

The theoretical sources offer no statements on this topic, nor can any unambiguous inference be drawn.

2. *Blacky Test*

a. Oral Sadism vs. Anal Retentiveness

Males: $r_t = -.40$, $P < .01$

Those males who are strongly oral-sadistic tend to give fewer strong anal-retentive responses.

3. *Interpretation*

The negative correlation obtained between Oral Sadism and Anal Retentiveness may result indirectly from an artifact in the relationship between Anal Expulsiveness and Anal Retentiveness. Scores on both anal dimensions are derived primarily from responses to Cartoon III, and choice of one type of response by the subject automatically decreases his likelihood of choosing the other. This relationship between the two anal dimensions is mutually exclusive in the Inquiry and tends to be inverse in the Spontaneous Story. Consequently, the negative correlation between Anal Expulsiveness and Anal Retentiveness ($r_t = -.38$, $P < .01$) is a function of the scoring system. Though the positive correlation between Oral Sadism and Anal Expulsiveness ($r_t = .25$) is not significant at the 5 per cent level, it may be sufficiently strong to entail a spurious negative correlation between Oral Sadism and Anal Retentiveness.[12]

E. Oral Sadism vs. Guilt Feelings (Females)

1. *Psychoanalytic Theory*

According to psychoanalytic theory, guilt feelings and orality are intimately connected. Fenichel elaborates further on this relationship:

[12]This artifact largely accounts for two other statistically significant inverse relationships with Anal Retentiveness (Oedipal Intensity for females; Sibling Rivalry for combined males and females). The latter two correlations will not be given separate treatment because of their probably spurious character.

"Guilt feeling not only has an oral character in general but an oral-sadistic character in particular" (1, p. 138).

The theory makes no differentiation between males and females in this respect.

2. *Blacky Test*

a. Oral Sadism vs. (Very Strong) Guilt Feelings.

Females: $r_t = .56$, $P < .01$

No correlation between oral sadism and guilt was found for the regular tetrachoric split of dimensional scores into Very Strong (+-+) plus Fairly Strong (+-) as one variable vs. Weak or Absent (0) as the other. However, when the dichotomy for Guilt Feelings was altered to Very Strong vs. Fairly Strong plus Weak or Absent, the correlation for the females jumped to significance at the 1 per cent level.[18] In other words, those females who have greater oral-sadistic conflicts are more likely to be among the Very Strong in guilt feelings.

3. *Interpretation*

The theoretical association between oral sadism and guilt is substantiated by the Blacky Test, but only in the case of females with scores of Very Strong on the dimension of Guilt Feelings.

F. Oral Sadism vs. Positive Ego Ideal (Combined Males and Females)

1. *Psychoanalytic Theory*

Theoretical inference must be called upon here as in the relationship between Oral Eroticism and Positive Ego Ideal in males. In Section 1 of Topic C, oral conflicts are said to interfere with the formation of a positive ego ideal in males. Since strong oral sadism is also an expression of oral disturbance, a similar inverse relationship with positive ego ideal is to be expected. In the case of females, a negative correlation can likewise be inferred, for the mother who is a frustrating object of aggression in the oral period is not likely to be introjected later as a highly respected, inspirational ideal.

---

[18] This departure from the routine method of analyzing the data weakens the argument. Such expedients capitalize on chance and reduce by unknown amounts the real significance of the correlations so obtained. No other similar departures from standard procedure occurred in this study.

2. *Blacky Test*

a. Oral Sadism vs. Positive Ego Ideal.

(Males: $r_t = -.45$)
(Females: $r_t = -.32$)
Combined: $r_t = -.39$, $P < .01$

Those subjects who score strongly on oral sadism are less likely to show evidence of having formed positive ego ideals.

3. *Interpretation*

The inverse correlation between oral sadism and positive ego ideal, inferred from psychoanalytic theory, is supported by the Blacky Test result.

G. Anal Expulsiveness vs. Masturbation Guilt (Combined Males and Females

1. *Psychoanalytic Theory*

Evidence on this topic can only be deduced in a highly circuitous and uncertain fashion. Masturbation guilt, according to Fenichel, is directly related to oedipal fantasies:

> . . . In the phallic phase, these fantasies, as a rule, express more or less directly the Oedipus complex. In adolescence and later life, frequently not only fears and guilt feelings are still connected with masturbation but there is even a distinct resistance on the part of the patients against enlightenment about the harmless nature of masturbation. . . . Analysis, as a rule, shows that a guilt feeling arising from the tendencies of the Oedipus complex has been displaced toward the activity that serves as an outlet for these unconscious fantasies (the conscious masturbatory fantasies being a distorted derivative of the unconscious Oedipus fantasies) . . . (1, pp. 75-6).

Therefore, since masturbation guilt is largely a function of oedipal intensity which, in turn, is supposed to be positively correlated with anal expulsiveness (see Topic A), a positive relationship might conceivably be inferred between anal expulsiveness and masturbation guilt.

2. *Blacky Test*

a. Anal Expulsiveness vs. Masturbation Guilt.

(Males: $r_t = .25$)
(Females: $r_t = .20$)
Combined: $r_t = .23$, $P < .05$

Those who reveal stronger anal-expulsive features tend to show greater masturbation guilt.

3. *Interpretation*

If the dubious connection established in Section 1 between anal expulsiveness and masturbation guilt is legitimate, then the test finding is consistent with it.

H. Anal Expulsiveness vs. Castration Anxiety (Males)

1. *Psychoanalytic Theory*

The theory is quite explicit in postulating a close relationship between anality and castration anxiety. Quotations from Fenichel on this topic are:

> Again it is possible that genital Oedipus wishes have been regressively replaced by anal impulses, and that in this way castration anxiety has regressively been supplanted by anal fears. The frequent bathroom and toilet phobias which are observed in children and in compulsion neurotics, such as a fear of falling into the toilet, of being eaten up by some monster coming from it, or the rationalized fear of being infected, as a rule show signs of a condensation of ideas of dirt with castration ideas (1, p. 200).
>
> As in oral fears, the fact that anal fears cover castration anxieties does not contradict the autonomous nature of pregenital fears. This distortion of castration anxiety is a *regressive* one, formed by the remobilization of the old pregenital anxiety over the loss of feces. It is often very difficult to determine which fraction of an anal anxiety represents a vestige of original pregenital anxiety, perhaps contributing a certain quality to the castration fear right from the beginning (the pregenital experiences of parting with the breast and with the fecal masses are archaic forerunners of the idea of castration), and which fraction is regressively distorted castration anxiety (1, p. 276).

2. *Blacky Test*

a. Anal Expulsiveness vs. Castration Anxiety.

Males: $r_i = .36$, $P < .02$

Those males who show greater disturbance in anality tend to exhibit more signs of castration anxiety.

3. *Interpretation*

Theory and test agree specifically in regard to the positive association between anality and castration anxiety.

I. Anal Expulsiveness vs. Sibling Rivalry (Combined Males and Females)

1. *Psychoanalytic Theory*

Freud and Fenichel do not offer any definitive formulation in this connection, but the postulation of a positive relationship follows readily from the common root of the two dimensions in demands not satisfied by the parents. Anal expulsiveness is an expression of hostility toward the parents in retaliation for frustrations imposed by the latter. Sibling rivalry is an expression of hostility toward a sibling in retaliation for diverting parental love, which is considered to be an important source of frustration.

2. *Blacky Test*

a. Anal Expulsiveness vs. Sibling Rivalry

(Males: $r_t = .23$)
(Females: $r_t = .20$
Combined: $r_t = .22$, $P < .05$

Those subjects who receive strong scores on the dimension of Anal Expulsiveness tend to show stronger evidence of sibling rivalry.

3. *Interpretation*

The association of anal expulsiveness and sibling rivalry inferred from the theory is borne out by the positive test correlation between the two dimensions. The test finding may possibly be attributable to a common, generalized aggression factor — those who are generally more aggressive will tend to express greater hostility toward parents and siblings alike. However, this again poses the hen-egg dilemma, for generalized aggression may arise from the specific relationships postulated by the theory. In any event, the Blacky Test data are not inconsistent with the inferred theoretical position on this topic.

J. Anal Expulsiveness vs. Positive Ego Ideal (Combined Males and Females)

1. *Psychoanalytic Theory*

Theoretical inference here follows from previous discussions of the positive ego ideal (see Topics *C*, *F*). Those who are preoccupied by anal expressions of hostility toward their parents are less likely to introject the parents as ego ideal figures.

2. *Blacky Test*

a. Anal Expulsiveness vs. Positive Ego Ideal.

(Males: $r_t = -.27$)
(Females: $r_t = -.25$)
Combined: $r_t = -.26$, $P < .05$

Those who show disturbances at the anal level tend not to adopt positive ego ideals based on parental introjections.

3. *Interpretation*

Theoretical inference and test result are in accord concerning the inverse relationship between anal expulsiveness and positive ego ideal.

K. Oedipal Intensity vs. Sibling Rivalry (Females)

1. *Psychoanalytic Theory*

Psychoanalytic theory maintains that disturbances occasioned by the presence of a sibling, expressed in hostility toward that sibling, are linked with conflicts of oedipal origin. Fenichel says:

> Another important typical traumatic event is the birth of a sibling. This may be experienced as a sudden disturbance of Oedipus gratifications because the mother's care must now be shared with somebody else. . . . (1, p. 92).

According to Freud:

> The next accusation against the mother flares up when the next child makes its appearance in the nursery. . . . It [the older female child] feels that it has been dethroned, robbed and had its rights invaded, and so it directs a feeling of jealous hatred against its little brother or sister, and develops resentment against its faithless mother. . . . (6, pp. 167-8).

2. *Blacky Test*

a. Oedipal Intensity vs. Sibling Rivalry.

Females: $r_t = .37$, $P < .05$

Those females who reveal more intense oedipal conflicts also exhibit stronger sibling rivalry.

3. *Interpretation*

The Blacky Test finding for females is consistent with the theoretical observation that oedipal intensity and sibling rivalry are positively associ-

ated. The test result may also be accounted for in terms of a generalized jealousy factor; i.e., those who tend to be more jealous of parents will likewise tend to be more jealous of brothers or sisters. However, this view, too, may owe its origin to relationships postulated by the theory.

L. Oedipal Intensity vs. Narcissistic Love-Object (Males)

1. *Psychoanalytic Theory*

According to psychoanalytic theory, males who choose narcissistic love-objects (representing characteristics of their own personality) are characterized by unresolved oedipal conflicts and tend to identify themselves with their mother rather than with their father. Fenichel states:

> The type of individual who is more narcissistic than "feminine" endeavors first of all to secure a substitute for his Oedipus strivings. Having identified himself with his mother, he behaves as he previously had wished his mother to behave toward him. . . . (1, p. 332).

Additional evidence of the theoretical view is based upon the intimate relationship postulated between narcissism and homosexuality (1). Concerning the latter, Fenichel reports:

> The majority of [male] homosexuals not only present an Oedipus love for their mothers, just as do neurotic individuals, but for the most part the intensity of the mother fixation is even more pronounced. . . . (1, p. 331).

Freud discusses this topic as follows:

> . . . First there is the fixation on the mother, which renders passing on to another woman difficult. The identification with the mother is an outcome of this attachment, and at the same time in a certain sense it enables the son to keep true to her, his first object. Then there is the inclination towards a narcissistic object-choice, which lies in every way nearer and is easier to put into effect than the move towards the other sex. . . . (3, p. 241).

2. *Blacky Test*

a. Oedipal Intensity vs. Narcissistic Love-Object.

Males: $r_t = .44$, $P < .01$

Those males who show greater oedipal disturbance are more likely to choose narcissistic love-objects.

3. *Interpretation*

The test data substantiate the theoretical expectation of a positive correlation between Oedipal Intensity and Narcissistic Love-Object in Males.

M. Positive Identification vs. Narcissistic Love-Object (Males)

1. *Psychoanalytic Theory*

This relationship has already been discussed in the preceding section (see Topic L). Narcissistic males are said to make negative identifications, that is, with mother instead of with father.

2. *Blacky Test*

a. Positive Identification vs. Narcissistic Love-Object.

Males: $r_t = -.46$, $P < .02$

Those males who make positive paternal identifications are less likely to choose narcissistic love-objects.

3. *Interpretation*

Theory and test are again in specific agreement.[14]

N. Penis Envy vs. Narcissistic Love-Object (Females)

1. *Psychoanalytic Theory*

The theory predicts a close relationship between penis envy and narcissism in females, since the lack of a penis supposedly is conceived by the little girl as a narcissistic deprivation. In Freud's words:

> . . . the little girl . . . finds her enjoyment of phallic sexuality spoilt by the influence of penis-envy. She is wounded in her self-love by the unfavourable comparison with the boy who is so much better equipped. . . . (6, p. 172).
>
> . . . Their [women's] vanity is partly a further effect of penis-envy, for they are driven to rate their physical charms more highly as a belated compensation for their original sexual inferiority. . . . (6, p. 180).

2. *Blacky Test*

a. Penis Envy vs. Narcissistic Love-Object.

Females: $r_t = .34$, $P < .05$

Those females who show greater evidence of penis envy tend to score more strongly on the dimension of Narcissistic Love-Object.

---

[14]Since narcissistic object-choice in males is supposed to be associated with a strong, unresolved Oedipus complex and absence of father identification *in the same individual*, a supplementary correlation was computed between Narcissistic Love-Object and combined Oedipal Intensity and Negative Identification: $r_t = .55$, $P < .01$.

3. *Interpretation*

The theoretical relationship between penis envy and narcissism in females is borne out by the Blacky Test finding.

O. Positive Identification vs. Sibling Rivalry (Females)

1. *Psychoanalytic Theory*

No direct evidence on this topic has been found in Freud or Fenichel. Nor can a clear inference be drawn, since the female can react in different ways to frustrations imposed by the mother, which are implicit in the sibling rivalry situation. One possibility is that she may make a positive identification with the mother through the mechanism of "identification with the aggressor." Another is that she may resort to a negative identification with the father in order to win over the frustrating mother.

2. *Blacky Test*

a. Positive Identification vs. Sibling Rivalry.

Females: $r_t = .36$, $P < .05$

Those females who make positive maternal identifications also show greater evidence of sibling rivalry.

3. *Interpretation*

Agreement between theory and test cannot be evaluated in this connection. The test data suggest the plausibility of the "identification with the aggressor" hypothesis to explain female identifications.

P. Sibling Rivalry vs. Anaclitic Love-Object (Females)

1. *Psychoanalytic Theory*

There is also no definitive theoretical position on whether those females who choose anaclitic love-objects (reminding them of their fathers) are less or more likely to show greater sibling rivalry.

2. *Blacky Test*

a. Sibling Rivalry vs. Anaclitic Love-Object.

Females: $r_t = .40$, $P < .02$

Those females who score more strongly on the dimension of Sibling Rivalry tend to choose love-objects who bear characteristics of their fathers.

3. *Interpretation*

No evaluation of the theory is possible on this topic, but the test data seem to be consistent with the finding of the previous section (Topic *O*). If females who have more sibling rivalry are more likely to identify with their mothers, then they are also more likely to choose a love-object resembling the one chosen by the mother, namely, the father (see Topic *Q* for further elaboration).

Q. Positive Ego Ideal vs. Anaclitic Love-Object (Combined Males and Females)

1. *Psychoanalytic Theory*

Theoretical inference can be invoked concerning this relationship. A positive ego ideal, as the term has been used in this study, implies the incorporation of wholesome standards of the same-sex parent. The boy considers his father to be a shining example and wishes to follow in his footsteps, while the girl maintains a similar attitude toward her mother. The transition from idealization process to choice of a love-object similar to the one chosen by the ideal follows naturally.

2. *Blacky Test*

a. Positive Ego Ideal vs. Anaclitic Love-Object.

(Males: $r_t = .61$)
(Females: $r_t = .38$)
Combined: $r_t = .50$, $P < .01$

Those subjects who have adopted positive ego ideals are more likely to choose anaclitic love-objects.

3. *Interpretation*

Theoretical inference and test results are in accord with respect to the positive correlation between the dimensions of Positive Ego Ideal and Anaclitic Love-Object.

R. Summary

Now to summarize the findings relevant to the syntax of the psychoanalytic theory of psychosexual development. Intercorrelations of all Blacky Test dimensions were computed for males alone, females alone, and combined males and females. According to the design of the experiment, each statistically significant ($P < .05$) intercorrelation was examined in items

of its agreement or disagreement with the theoretical position as stated by Freud and Fenichel. The latter sources furnished *specific* evidence on eight of the 17 intercorrelational topics (*A, B, E, H, K, L, M, N*); *inferential* evidence on six (*C, F, G, I, J, Q*); and *no* evidence on three (*D, O, P*). As in the case of sex differences, the theoretical sources are incomplete with respect to definitive information on many of the topics considered. But where the theory *does* offer evidence, the consonance is even more striking than in the sex differences chapter. Blacky Test findings and psychoanalytic theory agree in *all eight* relationships for which there are specific theoretical viewpoints, and in *all six* for which theoretical evidence has been inferred!

Thus, the psychoanalytic theory of psychosexual development has literally been "put to the test" both in areas of significant sex differences and syntactical relationships. The final chapter will attempt to answer the paramount question: In light of the evidence, what conclusions can justifiably be drawn?

VII. IN LIGHT OF THE EVIDENCE

Psychoanalysis carries a lot of emotional freight. For some it is a way of life which requires constant devotion. For others it is a mystical cult which should be relegated to the realm of faith-healing. Even those whose opinions lie between the extremes are seldom dispassionate in their voicings. So by now the Smug Sympathizer has in all probability seized upon the striking agreements between theory and test, and the Caustic Critic has contented himself with the exploratory nature of the study and the incompleteness of the theory. But for the benefit of the inquisitive young clinical psychologist of Chapter I, as well as for the still-untainted men of science, an appraisal of the current study will be undertaken.

The Blacky Test, a modified projective technique, was administered to 119 male and 90 female students in elementary psychology classes at Stanford University. The protocols were scored in the form of analogues of psychoanalytic dimensions of psychosexual development. Comparing responses of the two sexes revealed a number of statistically significant differences between them. Intercorrelating scores on all of the dimensions also produced a number of statistically significant relationships. In neither case could the results reasonably be attributed to the operation of chance factors alone. The significant test data, including sex differences and dimensional intercorrelations, fell into 31 areas. Evidence was then sought in the writings of Freud and Fenichel to determine whether the significant test findings were consistent with psychoanalytic theory or not. Agreement between theory and test was subsequently noted in 14 of the 15 test areas in which the viewpoint of the theory was found to be stated specifically, and in all eight in which a theoretical viewpoint could be inferred. In the remaining eight areas, the theory provided neither postulates nor material suitable for inference, so that no comparisons were possible.

Several questions immediately come to the fore: Does the approach used in this study represent an advance over previous studies of the sort described by Sears? What are its limitations? In view of the limitations, what contributions does it make?

The approach differs from most earlier efforts to investigate psychoanalytic theory. It seeks to evaluate the theory in proper context by utilizing the concepts as well as the framework of the theory. Insofar as it strives for this goal, it represents a distinct advance.

The limitations of exploratory research are necessarily broad. The Blacky Test in its current form is not a clinically validated instrument. It has only been *assumed* that the test analogues can be equated with their

theoretical counterparts. This assumption may later prove to be unwarranted. Secondly, the study has been restricted to statistically significant test findings. This has reduced the completeness of the criterion (Blacky Test) as well as the representativeness of the sample of theoretical assertions examined. Perhaps a future investigation, in which meaning can safely be attached to insignificant findings, will alter the picture considerably. Thirdly, generalizations have to be confined to the particular sample of students tested in the experiment. So the results of the study must therefore be regarded as *tentative* and *suggestive* rather than as *final* and *conclusive*.

In this light, however, a great many contributions offer themselves. The most apparent is the striking consonance between significant findings from the test and the corresponding postulates from the theory. According to the design of the experiment, the psychoanalytic theory of psychosexual development was free to agree or disagree with the statistically significant test data. The fact that specific agreement occurred in 14 of 15 cases is highly impressive. It cannot be written off simply as an act of chance. *These data strongly suggest that some aspects of psychoanalytic theory have demonstrable validity.*

Apart from suggestions concerning the validity of the theory, the test findings also cast some light on specific factors in the psychosexual development of males and females. These factors lend themselves to the following speculative interpretations, which can serve as hypotheses to be checked in future investigations. In both sexes the early stages of development appear to be interrelated to the extent that disturbances at one level are associated with disturbances at the others. In the case of males, castration anxiety, originating in the phallic period, is also accompanied by earlier oral and anal conflicts. The male typically resolves the Oedipus situation by identifying predominantly with a father-figure and by introjecting that figure as his superego. Those who depart from the normal sequence by later seeking narcissistic love-objects are found to have strong, unresolved Oedipus conflicts accompanied by the lack of a positive father identification. The growing influence of the mother in American families is reflected in the proportion of males (almost one-third) whose superegos do appear to contain more maternal than paternal features. Generally the superego in males, according to the test data, represents more a threat of external punishment than an internalized fear of losing love. The setting up of a positive ego ideal seems to vary inversely with early oral and anal fixations, and directly with anaclitic type of object-choice.

The psychosexual development of the female, as portrayed by the Blacky Test findings, is more involved. The early pre-oedipal ambivalence of the girl toward her mother seems to have a pronounced effect on the entire developmental process. Strong oral-sadistic tendencies persist along with repressed anal sadism. The oedipal involvement is less complete than in the case of males, since fear of losing the love of the frustrating mother continues as a dominant motif. The subsequent identification process is less clear-cut, and it appears that largely through the mechanism of "identification with the aggressor" the girl is able to pattern herself after the mother toward whom she still harbors strong undercurrents of hostility. Once the mother has been introjected as the superego, the aggressions formerly directed toward her are turned inward and result in strong guilt feelings. The disturbed sequence culminates in the greater incidence of narcissistic type of object-choice in females. Penis envy also seems to play a prominent rôle in the latter connection.

Another suggested contribution of the study concerns the incompleteness of the psychoanalytic theory of psychosexual development, as presented by Freud and Fenichel. Of the 31 test areas studied, theoretical postulates for only 15 were found to be stated specifically in the sources. Evidence for an additional eight test areas could be inferred, with varying degrees of confidence. No theoretical hypotheses pertinent to the remaining eight were uncovered. Do these facts dim the psychoanalytic horizon? Many may argue that they do not, on the grounds that the theory was not designed to be tested in such a fashion and that its usefulness can never rest upon the findings of an artificial, almost indecent dissection. But there are those who hold that to enjoy the status of a theory it must submit to indefatigable scrutiny. In the eyes of the latter, the theory of psychosexual development, with particular reference to sex differences and syntactical relationships, requires sharpening and extending.

The reader may wish to raise a very legitimate question at this point: Where do we go from here? The current study, in the tradition of exploratory research, has raised more issues than it has settled. The data, though highly suggestive, do not furnish definitive answers. Can such answers ever be obtained? One hope lies in the future validation of the Blacky Test. If it can be demonstrated clinically that the test actually measures psychoanalytic dimensions, the data will increase in stature. All test results, statistically insignificant as well as significant, will acquire greater meaning. To accomplish such a validation is not an easy task. Some sort of plan could conceivably be carried out with the coöperation of a

group of psychoanalysts. If the Blacky Test were administered to each patient at the beginning of his analysis, the results written up (without being seen by the psychoanalyst) and put aside until such time as the analyst felt confident to offer his own judgments, a comparison could then be made of the two independent sources. Preliminary experiments in this general area have shown promise.

Every parable ends with a moral. This study, in which psychoanalytic theory paraded as a dog named Blacky, has three. It can caution the Smug Sympathizer to be aware of the theory's incompleteness. It can confront the Caustic Critic with results strikingly favorable to the theory. And it can console the young clinical psychologist with the possibility that the theory which he uses in his work may someday emerge from the scientific underworld.

APPENDICES

A. Appendix *A*: Blacky Test Inquiry for Males and Females

1. *For Males*

a. Cartoon I: Oral Eroticism.

(1). Is Blacky
  (*a*). happy?
  (*b*). unhappy?
  (*c*). or doesn't he feel one way or the other?

(2). How does Mama feel in this scene?
  (*a*). Very contented.
  (*b*). Pleased but tired.
  (*c*). Rather unhappy.

(3). Which would Blacky rather do?
  (*a*). Stay until his feeding is over and then go someplace else.
  (*b*). Stay as long as possible to be sure he gets enough nourishment.

(4). Which one of the following best describes Blacky?
  (*a*). He's a little glutton who never stops eating.
  (*b*). He's got a hearty appetite which usually gets satisfied.
  (*c*). He sometimes doesn't get enough to replace all the energy he burns up.

(5). Judging by appearances, how much longer will Blacky want to be nursed by Mama before being weaned?
  (*a*). He'll want to be on his own fairly soon.
  (*b*). He'll want to continue being nursed until he's quite a bit older.
  (*c*). He feels Mama would like to turn him loose right now.

(6). How will Blacky feel about eating when he grows older?
  (*a*). He will rather eat than do most anything else.
  (*b*). He will enjoy eating but will like lots of other things just as much.
  (*c*). He will never get enough to satisfy his appetite.

b. Cartoon II: Oral Sadism.

(1). Why is Blacky doing that to Mama's collar?

(2). How often does Blacky feel like acting up this way?
  (*a*). Once in a while.
  (*b*). Fairly often.
  (*c*). Very often.

(3). Blacky most often acts like this when he can't get enough of which one of the following?
  (*a*). Attention.
  (*b*). Milk.
  (*c*). Recreation.

(4). What will Blacky do next with Mama's collar?
  (*a*). Get tired of it and leave it on the ground?
  (*b*). Return it to Mama.
  (*c*). Angrily chew it to shreds.

(5). If Mama comes on the scene, what will she do?
  (*a*). Feed Blacky again.
  (*b*). Send him off to bed without his dinner.
  (*c*). Bark.
(6). What would Blacky do if Mama did come over to feed him?
  (*a*). He'd ignore Mama and continue chewing the collar.
  (*b*). He'd put down the collar and start eating.
  (*c*). He'd get even with Mama by trying to bite her instead of the collar.

c. Cartoon III: Anal Sadism.

(1). What was Blacky's main reason for defecating there?
  (*a*). He wanted to spite somebody. Who?
  (*b*). He was doing what Mama and Papa told him to.
  (*c*). He picked the spot by accident.
  (*d*). He wanted to keep his own area neat and clean.
(2). Which of the following is Blacky most concerned with here?
  (*a*). Throwing dirt over what he did so that it will be neatly covered up.
  (*b*). Relieving himself so that his system feels more comfortable.
  (*c*). Getting rid of his anger.
(3). Why is Blacky covering it up?
  (*a*). He wants to make as little mess as possible.
  (b). He doesn't want Mama and Papa to find out.
  (*c*). He's automatically doing what he's been taught.
(4). How does Blacky feel about the training he's been getting?
  (*a*). By relieving himself in the way he's been taught, he now has an opportunity to show his family what a good dog he can be.
  (*b*). He feels Mama and Papa are expecting too much of him at this early stage.
  (*c*). He is very happy to have control of himself.
  (*d*). He thinks he's got Mama and Papa right where he wants them.
(5). What will Mama say to Blacky?
(6). What will Papa say to Blacky?

d. Cartoon IV: Oedipal Intensity.

(1). How does Blacky feel about seeing Mama and Papa make love? Why?
(2). When does Blacky get this feeling?
  (*a*). Whenever he sees Mama or Papa.
  (*b*). Whenever he sees Mama and Papa together.
  (*c*). Whenever he sees Mama and Papa making love.
(3). Which one of the following makes Blacky most unhappy?
  (*a*). Papa keeping Mama all to himself.
  (*b*). The idea that Mama and Papa seem to be ignoring him on purpose.
  (*c*). He is ashamed watching them make love out in the open.
(4). What does Blacky suspect is the reason behind the scene he's watching?
  (*a*). He suspects Mama and Papa are planning an addition to the family.

(*b*). He suspects Mama and Papa are very much in love.
(*c*). He suspects Papa is having his own way about things.
(*d*). He suspects Mama and Papa are purposely depriving him of attention.

(5). What will Papa do if he sees Blacky pecking?
(6). What will Mama do if she sees Blacky pecking?
(7). Which would make a happier picture?
(*a*). Mama left on the outside watching Blacky together with Papa. Why?
(b). Papa left on the outside watching Blacky together with Mama. Why?

e. Cartoon V: Masturbation Guilt.

(1). How does Blacky feel here?
(*a*). Happy, without a care in the world.
(*b*). Enjoying himself, but a little worried.
(*c*). Mixed up and guilty.
(2). How might Blacky feel about this situation when he is older?
(*a*). Happy, without a care in the world.
(*b*). Enjoying himself, but a little worried.
(*c*). Mixed up and guilty.
(*d*). The situation won't come up again when he is older.
(3). Whom might Blacky be thinking about here?
(4). Does Blacky naively fear that something might happen to him? What?
(5). What will Mama say if she comes over and finds Blacky?
(6). What will Papa say if he comes over and finds Blacky?

f. Cartoon VI: Castration Anxiety.

(1). How does Blacky feel here?
(*a*). Terrified that he's going to be next.
(*b*). Puzzled and upset.
(*c*). Curious but calm.
(2). What does Blacky suspect might be the reason for this scene?
(*a*). He suspects Tippy is being punished for having done something wrong.
(*b*). He suspects Tippy is an innocent victim of someone else's ideas.
(*c*). He suspects Tippy is being improved in some way.
(3). How does Blacky feel about his own tail?
(*a*). He's not particularly worried about it.
(*b*). He's thinking desperately about a way to save it.
(*c*). He thinks he might look better if it is cut off.
(*d*). He's so upset he wishes he never saw or heard of tails.
(4). Do you suppose Blacky would prefer to have his own tail cut off right away rather than go through the suspense of wondering if it will happen to him? Why?

(5). Which member of the family most likely arranged for Tippy's tail to be cut off?
(6). What will other dogs in the neighborhood do when they see Tippy's short tail?
(*a*). Start worrying about their own tails
(*b*). Make fun of Tippy
(*c*). Wonder what's going on
(*d*). Admire Tippy

g. Cartoon VII: Positive Identification.
(1). Who talks like that to Blacky — Mama or Papa or Tippy?
(2). Whom is Blacky most likely to obey — Mama or Papa or Tippy?
(3). Whom is Blacky imitating here — Mama or Papa or Tippy?
(4). Whom would Blacky rather pattern himself after—Mama or Papa or Tippy?
(5). Blacky's disposition, actually, is most like the disposition of which one — Mama or Papa or Tippy?
(6). What would Blacky have an impulse to do if he were in the position of the toy dog?
(*a*). Get frightened and hide.
(*b*). Stand there and take it.
(*c*). Get mad and sulk.
(*d*). Start fighting.

h. Cartoon VIII: Sibling Rivalry.
(1). What does Blacky probably feel like doing now?
(*a*.) Beat Tippy up
(*b*). Bark happily at the group and join them
(*c*). Show up Tippy by doing something better
(*d*). Run away to spite Mama and Papa
(2). According to Blacky, how much praise does Tippy actually deserve?
(*a*). He feels Tippy fully deserves the praise.
(*b*). He feels Tippy deserves some praise, but not that much.
(*c*). He feels Tippy deserves to be punished instead of praised.
(3). Whom does Blacky feel is paying more attention to Tippy?
(*a*). Mama.
(*b*). Papa.
(*c*). Both paying the same amount.
(4). How often does Blacky see this?
(*a*). Once in a while.
(*b*). Fairly often.
(*c*). Very often.
(5). How does Blacky think Mama and Papa really feel toward him at this time?
(*a*). He thinks they love him more than they do Tippy.
(*b*). He thinks they love him about the same as they do Tippy.
(*c*). He thinks they love him less than they do Tippy.

(6). If Blacky is angry, whom is he most angry at — Mama or Papa or Tippy? Why?

i. Cartoon IX: Guilt Feelings.

(1). What might have happened between the last picture and this one?
(2). How is Blacky's conscience here?
  (*a*). His conscience is so strong he's practically paralyzed.
  (*b*). His conscience is bothering him somewhat, but he's mostly afraid of what will be done to him.
  (*c*). He's hardly bothered at all by his conscience, just afraid of what will be done to him.
(3). Which character do the actions of the pointing figure remind Blacky of?
(4). Who is really to blame for Blacky's feeling this way?
  (*a*). Himself
  (*b*). Somebody else — Who?
  (*c*). The situation couldn't be helped.
(5). How guilty does Blacky feel here?
  (*a*). He feels very guilty.
  (*b*). He feels fairly guilty.
  (*c*). He hardly feels guilty at all.
(6). What might Blacky do now?
(7). Do you think Blacky will
  (*a*). have this feeling as long as he lives?
  (*b*). feel badly every now and then?
  (*c*). feel badly for a little while and then go out to play?

j. Cartoon X: Postive Ego Ideal.

(1). Whom does the figure remind Blacky of?
(2). In Blacky's mind, how does Papa stack up against the dream figure when he compares them?
(3). What would be the main reason for Blacky wanting to be like the figure in his dream?
  (*a*). Then he would show up Tippy.
  (*b*). Then he would be the envy of all male dogs
  (*c*). Then he would be loved more by Mama and Papa.
  (*d*). Then he would be very popular with the females.
(4). What does Blacky himself probably feel about his chances of growing up to be like the figure in his dream?
  (*a*). He probably feels he has a very good chance to grow up to be like that.
  (*b*). He probably feels he has a fair chance to grow up to be like that.
  (*c*). He probably feels he has a very poor chance to grow up to be like that.
(5). Actually, what are Blacky's chances of growing up to be like the figure in his dream?

(*a*). Very good.
(*b*). Fair.
(*c*). Very poor.

(6). How often does Blacky probably have this kind of dream?
(*a*). Very often.
(*b*). Fairly often.
(*c*). Once in a while.

k. Cartoon XI: Love-Object.

(1). Who is the figure Blacky is dreaming about?
(2). Whom does the figure remind Blacky of?
(3). Which of the following possibilities would attract Blacky most?
(*a*). The possibility that the dream figure looks like himself, which would increase his pride.
(*b*). The possibility that the dream figure looks like Mama, which would remind him of the good old days.
(*c*). The possibility that the dream figure looks like someone else, whom he would make happy by giving her all his love.
(4). Why does Blacky feel so contented while he is dreaming?
(*a*). He feels everyone will admire him.
(*b*). He feels Mama will comfort him.
(*c*). He feels the dream figure will be delighted by his attentions.
(5). In Blacky's mind, how does Mama stack up against the dream figure when he compares them?
(6). Would Blacky rather be like the figure in his dream? Why?

2. *For Females*

a. Cartoon I: Oral Eroticism.

(1). Is Blacky
(*a*). happy?
(*b*). unhappy?
(*c*). or doesn't she feel one way or the other?
(2). How does Mama feel in this scene?
(*a*). Very contented.
(*b*). Pleased but tired.
(*c*). Rather unhappy
(3). Which would Blacky rather do?
(*a*). Stay until her feeding is over and then go someplace else.
(*b*). Stay as long as possible to be sure she gets enough nourishment.
(4). Which one of the following best describes Blacky?
(*a*). She's a little glutton who never stops eating.
(*b*). She's got a hearty appetite which usually gets satisfied.
(*c*). She sometimes doesn't get enough to replace all the energy she burns up.
(5). Judging by appearances, how much longer will Blacky want to be nursed by Mama before being weaned?

(*a*). She'll want to be on her own fairly soon.
(*b*). She'll want to continue being nursed until she's quite a bit older.
(*c*). She feels Mama would like to turn her loose right now.

(6). How will Blacky feel about eating when she grows older?
(*a*). She will rather eat than do most anything else.
(*b*). She will enjoy eating but will like lots of other things just as much.
(*c*). She will never get enough to satisfy her appetite.

b. Cartoon II: Oral Sadism.

(1). Why is Blacky doing that to Mama's collar?

(2). How often does Blacky feel like acting up this way?
(*a*). Once in a while.
(*b*). Fairly often.
(*c*). Very often.

(3). Blacky most often acts like this when she can't get enough of which one of the following?
(*a*). Attention.
(*b*). Milk.
(*c*). Recreation.

(4). What will Blacky do next with Mama's collar?
(*a*). Get tired of it and leave it on the ground.
(*b*). Return it to Mama.
(*c*). Angrily chew it to shreds.

(5). If Mama comes on the scene, what will she do?
(*a*). Feed Blacky again.
(*b*). Send her off to bed without her dinner.
(*c*). Bark.

(6). What would Blacky do if Mama did come over to feed her?
(*a*). She'd ignore Mama and continue chewing the collar.
(*b*). She'd put down the collar and start eating.
(*c*). She'd get even with Mama by trying to bite her instead of the collar.

c. Cartoon III: Anal Sadism.

(1). What was Blacky's main reason for defecating there?
(*a*). She wanted to spite somebody. Who?
(*b*). She was doing what Mama and Papa told her to.
(*c*). She picked the spot by accident.
(*d*). She wanted to keep her own area neat and clean.

(2). Which one of the following is Blacky most concerned with here?
(*a*). Throwing dirt over what she did so that it will be neatly covered up.
(*b*). Relieving herself so that her system feels more comfortable.
(*c*). Getting rid of her anger.

(3). Why is Blacky covering it up?
(*a*). She wants to make as little mess as possible.

(*b*). She doesn't want Mama and Papa to find out.
(*c*). She's automatically doing what she's been taught.
(4). How does Blacky feel about the training she's been getting?
(*a*). By relieving herself in the way she's been taught, she now has an opportunity to show her family what a good dog she can be.
(*b*). She feels Mama and Papa are expecting too much of her at this early stage.
(*c*). She is very happy to have control of herself.
(*d*). She thinks she's got Mama and Papa right where she wants them.
(5). What will Mama say to Blacky?
(6). What will Papa say to Blacky?

d. Cartoon IV: Oedipal Intensity.

(1). How does Blacky feel about seeing Mama and Papa make love? Why?
(2). When does Blacky get this feeling?
(*a*). Whenever she sees Mama or Papa.
(*b*). Whenever she sees Mama and Papa together.
(*c*). Whenever she sees Mama and Papa making love.
(3). Which one of the following makes Blacky most unhappy?
(*a*). Mama keeping Papa all to herself.
(*b*). The idea that Mama and Papa seem to be ignoring her on purpose.
(*c*). She is ashamed watching them make love out in the open.
(4). What does Blacky suspect is the reason behind the scene she's watching?
(*a*). She suspects Mama and Papa are planning an addition to the family.
(*b*). She suspects Mama and Papa are very much in love.
(*c*). She suspects Papa is having his own way about things.
(*d*). She suspects Mama and Papa are purposely depriving her of attention.
(5). What will Papa do if he sees Blacky peeking?
(6). What will Mama do if she sees Blacky peeking?
(7). Which would make a happier picture?
(*a*). Mama left on the outside watching Blacky together with Papa. Why?
(*b*). Papa left on the outside watching Blacky together with Mama. Why?

e. Cartoon V: Masturbation Guilt.

(1). How does Blacky feel here?
(*a*). Happy, without a care in the world.
(*b*). Enjoying herself, but a little worried.
(*c*). Mixed up and guilty.
(2). How might Blacky feel about this situation when she is older?
(*a*). Happy, without a care in the world.

(*b*). Enjoying herself, but a little worried.
(*c*). Mixed up and guilty.
(*d*). The situation won't come up again when she is older.
(3). Whom might Blacky be thinking about here?
(4). Does Blacky naively fear that something might happen to her? What?
(5). What will Mama say if she comes over and finds Blacky?
(6). What will Papa say if he comes over and finds Blacky?

f. Cartoon VI: Penis Envy.

(1). How does Blacky feel about her own tail?
(*a*). She has resigned herself to the inevitable.
(*b*). She's thinking desperately about a way to save it.
(*c*). She thinks she might look better if it is cut off.
(2). What would Blacky be most upset about if she were in Tippy's place?
(*a*). The fact that nobody loved her enough to prevent this from happening.
(*b*). The fact that she would not have her tail any more.
(*c*). The fact that she had allowed herself to be bad enough to deserve this.
(3). Which member of the family most likely arranged for Tippy's tail to be cut off?
(4). How will Tippy feel afterward about having had the tail cut off?
(*a*). Tippy will always be envious of dogs that have tails to wag.
(*b*). Tippy will try to make the best of a bad situation.
(*c*). Tippy will be proud to be different from the others.
(5). What will other dogs in the neighborhood do when they see Tippy's short tail?
(*a*). Start worrying about their own tails.
(*b*). Make fun of Tippy.
(*c*). Wonder what's going on.
(*d*). Admire Tippy.
(6). How would Blacky feel about trading her tail for a pretty bow which the male dogs would all admire?

g. Cartoon VII: Positive Identification.

(1). Who talks like that to Blacky — Mama, Papa, or Tippy?
(2). Whom is Blacky most likely to obey — Mama, Papa, or Tippy?
(3). Whom is Blacky imitating here — Mama, Papa, or Tippy?
(4). Whom would Blacky rather pattern herself after — Mama, Papa, or Tippy?
(5). Blacky's disposition actually, is most like the disposition of which one — Mama, Papa, or Tippy?
(6). What would Blacky have an impulse to do if she were in the position of the toy dog?
(*a*). Get frightened and hide.

(*b*). Stand there and take it.
(*c*). Get mad and sulk.
(*d*). Start fighting.

h. Cartoon VIII: Sibling Rivalry.
(1). What does Blacky probably feel like doing now?
(*a*). Beat Tippy up.
(*b*). Bark happily at the group and join them.
(*c*). Show up Tippy by doing something better.
(*d*). Run away to spite Mama and Papa.
(2). According to Blacky, how much praise does Tippy actually deserve?
(*a*). She feels Tippy fully deserves the praise.
(*b*). She feels Tippy deserves some praise, but not that much.
(*c*). She feels Tippy deserves to be punished instead of praised.
(3). Whom does Blacky feel is paying more attention to Tippy?
(*a*). Mama.
(*b*). Papa.
(*c*). Both paying the same amount.
(4). How often does Blacky see this?
(*a*). Once in a while.
(*b*). Fairly often.
(*c*). Very often.
(5). How does Blacky think Mama and Papa really feel toward her at this time?
(*a*). She thinks they love her more than they do Tippy.
(*b*). She thinks they love her about the same as they do Tippy.
(*c*). She thinks they love her less than they do Tippy.
(6). If Blacky is angry, whom is she most angry at — Mama, Papa, or Tippy? Why?

i. Cartoon IX: Guilt Feelings.
(1). What might have happened between the last picture and this one?
(2). How is Blacky's conscience here?
(*a*). Her conscience is so strong she's practically paralyzed.
(*b*). Her conscience is bothering her somewhat, but she's mostly afraid of what will be done to her.
(*c*). She's hardly bothered at all by her conscience, just afraid of what will be done to her.
(3). Which character do the actions of the pointing figure remind Blacky of?
(4). Who is really to blame for Black'y feeling this way?
(*a*). Herself
(*b*). Somebody else — Who?
(*c*). The situation couldn't be helped.
(5). How guilty does Blacky feel here?
(*a*). She feels very guilty.
(*b*). She feels fairly guilty.
(*c*). She hardly feels guilty at all

(6). What might Blacky do now?
(7). Do you think Blacky will
  (*a*). have this feeling as long as she lives?
  (*b*). feel badly every now and then?
  (*c*) feel badly for a little while and then go out to play?

j. Cartoon X (Figure 12): Positive Ego Ideal.

(1). Whom does the figure remind Blacky of?
(2). In Blacky's mind, how does Mama stack up against the dream figure when she compares them?
(3). What would be the main reason for Blacky wanting to be like the figure in her dream?
  (*a*). Then she would show up Tippy.
  (*b*). Then she would be the envy of all female dogs.
  (*c*). Then she would be loved more by Mama and Papa.
  (*d*). Then she would be very popular with the males.
(4). What does Blacky herself probably feel about her chances of growing up to be like the figure in her dream?
  (*a*). She probably feels she has a very good chance to grow up to be like that.
  (*b*). She probably feels she has a fair chance to grow up to be like that.
  (*c*). She probably feels she has a very poor chance to grow up to be like that.
(5). Actually, what are Blacky's chances of growing up to be like the figure in her dream?
  (*a*). Very good.
  (*b*). Fair.
  (*c*). Very poor.
(6). How often does Blacky probably have this kind of dream?
  (*a*). Very often.
  (*b*). Fairly often.
  (*c*). Once in a while.

k. Cartoon XI (Figure 11): Love-Object.

(1). Who is the figure in Blacky's dream?
(2). Whom does the figure remind Blacky of?
(3). Which of the following possibilities would attract Blacky most?
  (*a*). The possibility that the dream figure looks like herself, which would increase her pride.
  (*b*). The possibility that the dream figure looks like her parents, which would remind her of the good old days.
  (*c*). The possibility that the dream figure looks like someone else, whom she would make happy by giving him all her love.
(4). Why does Blacky feel so contented while she is dreaming?
  (*a*). She feels everyone will admire her.
  (*b*). She feels contented thinking about Mama and Papa.
  (*c*). She feels the dream figure will be delighted by her attentions.

(5). In Blacky's mind, how does Papa stack up against the dream figure when she compares them?
(6). Would Blacky rather be like the figure in her dream? Why?

B. Appendix B: Scoring the Test Inquiry....

(Same for Both Sexes except Where Noted)

1. *Cartoon I: Oral Eroticism*

Subject is scored Strong if he picks two or more of the following:

Item 4 — (1)
" 5 — (3)
" 6 — (1) or (3)

2. *Cartoon II: Oral Sadism*

Subject is scored Strong if he picks two or more of the following:

Item 2 — (3)
" 3 — (2)
" 4 — (3)
" 6 — (3)

3. *Cartoon III: Anal Expulsiveness*

Each of the following is scored Strong and included separately in the total dimensional score:

Item 1 — (1)
" 2 — (3)
" 3 — (2)
" 4 — (4)
" 5 & 6 — Qualitative judgment

4. *Cartoon III: Anal Retentiveness*

Each of the following is scored Strong and included separately in the total dimensional score:

Item 1 — (4)
" 2 — (1)
" 3 — (1)
" 4 — (1)
" 5 & 6 — Qualitative judgment

5. *Cartoon IV: Oedipal Intensity*

Subject is scored Strong if he picks two or more of the following:

Item 2 — (1) or (2)
" 3 — (1)
" 4 — (3)
" 7 — Qualitative judgment

6. *Cartoon V: Masturbation Guilt*

Subject is scored Strong if he picks two or more of the following:

Item 1 — (2) or (3)
" 2 — (2) or (3) or (4)
" 4 — Qualitative judgment

7. *Cartoon VI: Castration Anxiety* (Males)

Subject is scored Strong if he gets score of two or more on the following:

Item 1 — (1) gets wt. of 2 or (2) gets wt. of 1.
" 3 — (2) or (4) each get wt. of 2
" 4 — Qualitative judgment (wt. of 2 or 1)
" 6 — (1) gets wt. of 2 or (2) gets wt. of 1

8. *Cartoon VI: Penis Envy (Females)*

Subject is scored Strong if she gets score of three or more on the following:

Item 2 — (2) gets wt. of 1
" 4 — (1) gets wt. of 2
" 5 — (2) gets wt. of 2
" 6 — Qualitative judgment (wt. of 1)

9. *Cartoon VII: Positive Identification*

Male subject is scored Strong if he answers four or more of the following as "Papa":

Item 1
" 2
" 3
" 4
" 5

Female subject is scored Strong if she answers four or more of the above as "Mama."

10. *Cartoon VIII: Sibling Rivalry*

Subject is scored Strong if he picks one or more of the following:

Item 1 — (1) or (4)
" 2 — (3)
" 4 — (3)

11. *Cartoon IX: Guilt Feelings*

Subject is scored Strong if he picks two or more of the following:

Item 2 — (1)
" 5 — (1)
" 7 — (1) or (2)

12. *Cartoon X: Positive Ego Ideal*

Items 1 and 2 are scored qualitatively and included separately in the total dimensional score.

13. *Cartoon XI: Narcissistic Love-Object*

Each of the following is scored Strong and included separately in the total dimensional score.

Item 1 — "Blacky"
" 2 — "Blacky"
" 3 — (1)
" 4 — (1)
" 6 — Qualitative judgment

14. *Cartoon XI: Anaclitic Love-Object*

Each of the following is scored Strong and included separately in the total dimensional score:

Item 1 — "Mama" for males; "Papa" for females
" 2 — "Mama" for males; "Papa" for females
" 3 — (2)
" 4 — (2)
" 5 — Qualitative judgment

C. Appendix C. Standards for Scoring Spontaneous Stories

The following are the standards for scoring Spontaneous Stories on the various cartoons as "Strong." All others are scored "Not Strong."

1. *Cartoon I: Oral Eroticism*

Complete evasion of feeding reference.
B stealing milk while *M* asleep.
B too large to be nursing.
Mention of extreme deprivation — "starving to death," "*M* reluctant to feed *B*."
Very uncomfortable eating position — "unnatural."
Abnormal elements — "*B* biting *M*," "*B* asleep while eating," "*B* is adopted child."
Denial or unwillingness to describe emotion — "can't tell how *B* feels",
B will never be weaned — "spoiled," "always hungry."
B very persistent — "won't be denied".
Strong emotion re feeding — "oblivious," "delirious," "getting his fill."
Physical description of *B*'s satisfaction — "crouched to get all he can," "tail wagging vigorously."
Feeding references concerning *P* or *T* (more than just looking for a bone).
Freudian slips—"except" instead of "accept".

2. *Cartoon II: Oral Sadism* (including Item 1 of Inquiry)

Specific denial of aggression — "*B* not angry at *M*, just playing."
Emphasis on chewing — "likes to chew all the time."
Any mention of killing.
Any act or threat of violence toward *M*, regardless of by whom — "*M* run over by car."
Mention of feeding deprivation (more than "just finished eating").

3. *Cartoon III: Anal Expulsiveness*

B angry, spiteful (wt. of 2).
B triumphant, very happy (wt. of 2)
B somewhat annoyed (wt. of 1).
B feels better when relieved (wt. of 1).
B ate too much (wt. of 1).
(Above are additive up to score of 3).

4. *Cartoon III: Anal Retentiveness*

B strongly objects to dirtiness of smell (wt. of 2).
B wants to please parents (wt. of 2).
Any mention of "treasure" (wt. of 2).
B mildly objects to dirtiness or smell (wt. of 1).
Any mention of covering, hiding, burying, or "making a deposit" (wt. of 1).
(Above are additive up to a score of 3).

5. *Cartoon IV: Oedipal Intensity*

Specific jealousy of parents' love for each other, not primarily due to lack of attention.
Specifically wanting parents apart, wanting to disturb them.
Hostility toward either *M* or *P*, not primarily due to lack of attention.
Denial of *B*'s emotional involvement — "disinterested in proceedings."
Strong shame or guilt (may be displaced on to others) — "sneaking," "spying," "shameful."
Deriding males or females as such — "girls are no good."

6. *Cartoon V: Masturbation Guilt*

Complete avoidance of sex or masturbation.
Denials — "no significance," "not worried," "not discovering sex".
Emphasis on normality — "all dogs do it."
Fear of being discovered.
Emphasis on *B*'s unhappiness.
Other strong evidences of guilt — "perverted action," "ashamed," "afraid to ask parents."

7. *Cartoon VI: Castration Anxiety* (Males)

Any mention of *B*'s own tail — "thinks his tail is next," "own tail sticking out."
Denial of *B*'s involvement — "not concerned, just curious."
Special concern over any tail — "*T* has such a long, wonderful tail."
Unusual resistance — "no point to story at all".

8. *Cartoon VI: Penis Envy* (Females)

Remarks concerning tail per se, admiring it, wish to save own tail.
Resentment of notion that tail is being removed for a good purpose.
Regret about inevitability of tail being cut off.
Specific satisfaction at male *T*'s tail being cut off.

9. *Cartoon VII: Positive Identification*

Specific mention of *B* imitating same-sex parent.

10. *Cartoon VIII: Sibling Rivalry*

Specific mention of hatred for *T*.
B wants vengeance, wants to break up gathering.
B specifically scornful of *T*.
Expression of very strong emotion (more than sulking)—"wants to cry."
B thinking of running away from home.
B has harmed *T* or is contemplating doing so.
Obviously a forced happy story.
B is bored, thinks it's silly.
Denial of concern or unhappiness—"doesn't care."
Strongly evasive, avoids involvement.
Statement that *B* likes *T*, glad *T* is happy.
Freudian slips—"*T*" instead of "*B*."

11. *Cartoon IX: Guilt Feelings* (including Item 1 of Inquiry)

Complete avoidance of any reference to guilt.
Denial of conscience or shame—"only a dream."
Strong expressions of guilt—"afraid of going to Hell," "will feel like this forever after."
Any mention of death or killing.

12. *Cartoon X: Positive Ego Ideal*

Specifically relating *B*'s dream to same-sex parent.

13. *Cartoon XI: Narcissistic Love-Object*

Dream figure is *B* himself (herself).
Any mention of resemblance of dream figure to *B* himself (herself).

14. *Cartoon XI: Anaclitic Love-Object*

Dream figure is opposite-sex parent.
Dream figure is linked in some way to opposite-sex parent.

D. Appendix *D*: Derivation of Dimensional Scores

(Both Sexes)

1. *Oral Eroticism*

Each of the following scored as Strong or Not Strong:

a. Spontaneous Story (I).
b. Inquiry (I).
c. Preference (I).
d. Related Comments.

Strong on 3 or 4 out of 4 is given dimensional score of Very Strong (++).
Strong on 2 out of 4 is given dimensional score of Fairly Strong (+).
Strong on 0 or 1 out of 4 is given dimensional score of Weak or Absent (0).

2. *Oral Sadism*

Same as above, except based on Cartoon II.

3. *Anal Expulsiveness*

Each of the following scored as Strong or Not Strong:

a. Spontaneous Story (III).
b. Each Individual Item of the Inquiry (III).
c. Preference (III).
d. Related Comments.

Strong on 3 or more = Very Strong.
Strong on 2 = Fairly Strong
Strong on 0 or 1 = Weak or Absent.

4. *Anal Retentiveness*

Each of following scored as Strong or Not Strong:

a. Spontaneous Story (III).
b. Each Individual Item of the Inquiry (III).
c. Preference (III).
d. Related Comments.

Strong on 4 or more = Very Strong.
Strong on 3 = Fairly Strong.
Strong on 0, 1, 2 = Weak or Absent.

5. *Oedipal Intensity*

Same as Oral Eroticism, except based on Cartoon IV.

6. *Masturbation Guilt*

Same as Oral Eroticism, except based on Cartoon V.

7. *Castration Anxiety* (Males only)

Same as Oral Eroticism, except based on Cartoon VI.

8. *Penis Envy* (Females only)

Same as Oral Eroticism, except based on Cartoon VI.

9. *Positive Identification*

Each of following scored as Strong or Not Strong:

a. Spontaneous Story (VII).
b. Inquiry (VII).
c. Related Comments.
d. Item 3 on Cartoon IX Inquiry.

Strong on 3 or 4 = Very Strong.
Strong on 2 = Fairly Strong.
Strong on 0 or 1 = Weak or Absent.

10. *Sibling Rivalry*

Same as Oral Eroticism, except based on Cartoon VIII.

11. *Guilt Feelings*

Same as Oral Eroticism, except based on Cartoon IX.

12. *Positive Ego Ideal*

Each of following scored as Strong or Not Strong:
- *a.* Spontaneous Story (X).
- *b.* Item 1 on Cartoon X Inquiry.
- *c.* Item 2 on Cartoon X Inquiry.
- *d.* Related Comments.

Strong on any 2 out of + = Very Strong.
Strong on any 1 (except Item 2) = Fairly Strong.
Strong on 0 (or only Item 2) = Weak or Absent.

13. *Narcissistic Love-Object*

Each of following scored as Strong or Not Strong:
- *a.* Spontaneous Story (XI).
- *b.* Items 1, 2, 3, 4, 6 on Cartoon XI Inquiry.
- *c.* Related Comments (Double or single weight).

Strong on Spontaneous Story = Very Strong.
Strong on Item 6 or any 2 = Fairly Strong.
Strong on 0 or 1 = Weak or Absent.

14. *Anaclitic Love-Object*

Each of following scored as Strong or Not Strong:
- *a.* Spontaneous Story (XI).
- *b.* Items 1, 2, 3, 4, 5 on Cartoon XI Inquiry.
- *c.* Related Comments (Double or single weight).

Strong on Spontaneous Story = Very Strong.
Strong on any of Items 1-4 or Item 5 plus Related Comments or any 2 = Fairly Strong.
Strong on 0 or Item 5 alone = Weak or Absent.

E. Appendix *E*: Blacky Test Data Yielding No Statistically Significant Differences Between the Sexes

1. *Test Dimensions Yielding No Sex Difference:*
   - Oral Eroticism.
   - Oral Sadism.
   - Anal Expulsiveness.
   - Anal Retentiveness.
   - Masturbation Guilt.
   - Sibling Rivalry.
   - Positive Ego Ideal.
   - Anaclitic Love-Object.

2. *Items from the Inquiry Yielding No Sex Differences:*
   - I—1, 2, 3, 5, 6.
   - II—2, 3, 5, 6.
   - III—1, 2, 3.
   - IV—2.
   - VIII—1, 2, 3, 4, 5.
   - IX—5.
   - X—5.

3. *No sex difference in number of subjects ignoring introductory comment to Cartoon V* ("Here Blacky is discovering sex").

Oral Eroticism	.44*	.21	—.24	.29	—.10	.35*	—.11	.08	—.04	—.40*	.14	—.05
Oral Sadism		.25	—.40*	.36*	.10	—.04	—.09	.24	.10	—.45*	.06	—.10
Anal Expulsiveness			—.27	.17	.25	.56*	—.02	.23	—.04	—.27	.06	—.03
Anal Retentiveness				—.08	—.08	.12	—.10	—.20	.01	.19	.11	.09
Oedipal Intensity					—.03	—.08	.02	.15	.05	.15	.44*	.00
Masturbation Guilt						—.08	.00	.27	.01	.17	—.05	.08
Castration Anxiety							—.04	.09	.15	—.17	.10	.19
Positive Identification								—.09	—.09	—.09	—.46*	.03
Sibling Rivalry									.05	—.13	.23	—.01
Guilt Feelings										—.05	.14	.05
Positive Ego Ideal											.22	.61*
Narcissistic Love-Object												.10
Anaclitic Love-Object												

*Asterisks denote those relationships in which the chi-square P-values are $< .05$.

TABLE B

TABLE OF DIMENSIONAL INTERCORRELATIONS FOR FEMALES*

	Oral Eroticism	Oral Sadism	Anal Expulsiveness	Anal Retentiveness	Oedipal Intensity	Masturbation Guilt	Penis Envy	Positive Identification	Sibling Rivalry	Guilt Feelings	Positive Ego Ideal	Narcissistic Love-Object	Anaclitic Love-Object
Oral Eroticism		.30	.31	—.15	.26	.28	.01	—.23	—.14	.01	—.03	—.02	.05
Oral Sadism			—.04	.15	—.01	.17	—.01	—.22	—.21	—.03	—.32	—.19	—.10
Anal Expulsiveness				—.52*	.38*	.20	—.10	—.03	.20	.12	—.25	.17	.20
Anal Retentiveness					—.35*	.00	.24	.00	—.30	—.01	.18	.14	—.27
Oedipal Intensity						.17	—.01	—.08	.37*	.11	.00	—.05	.07
Masturbation Guilt							.08	.27	.10	—.32	—.25	—.09	—.10
Penis Envy								—.09	.07	—.07	.25	.34*	—.03
Positive Identification									.36*	.16	.10	.00	.08
Sibling Rivalry										—.08	.10	.18	.40*
Guilt Feelings											—.19	—.10	—.14
Positive Ego Ideal												.22	.38*
Narcissistic Love-Object													—.23
Anaclitic Love-Object													

*Asterisks denote those relationships in which the chi-square P-values are $< .05$.

TABLE C

TABLE OF DIMENSIONAL INTERCORRELATIONS FOR COMBINED MALES AND FEMALES*

	Oral Eroticism	Oral Sadism	Anal Expulsiveness	Anal Retentiveness	Oedipal Intensity	Masturbation Guilt	Positive Identification	Sibling Rivalry	Guilt Feelings	Positive Ego Ideal	Narcissistic Love-Object	Anaclitic Love-Object
Oral Eroticism		.17	.26		.28							
Oral Sadism										—.39		
Anal Expulsiveness				—.38		.23		.22		—.26		
Anal Retentiveness								—.25				
Oedipal Intensity												
Masturbation Guilt												
Positive Identification												
Sibling Rivalry												
Guilt Feelings												
Positive Ego Ideal												.50
Narcissistic Love-Object												
Anaclitic Love-Object												

*Only those relationships in which the chi-square *P*-values are $< .05$ are presented. Male and female data were combined only when similar trends were separately noted in each.

TABLE *D*
CARTOON TITLES

	Cartoon I	(Fig. 2)	Oral Eroticism
	" II	(Fig. 3)	Oral Sadism
	" III	(Fig. 4)	Anal Sadism
	" IV	(Fig. 5)	Oedipal Intensity
	" V	(Fig. 6)	Masturbation Guilt
	" VI	(Fig. 7)	Castration Anxiety (Males) Penis Envy (Females)
	" VII	(Fig. 8)	Positive Identification
	" VIII	(Fig. 9)	Sibling Rivalry
	" IX	(Fig. 10)	Guilt Feelings
(Males)	" X	(Fig. 11)	Positive Ego Ideal (Males)
(Females)	" XI	(Fig. 11)	Love-Object (Females)
(Males)	" XI	(Fig. 12)	Love-Object (Males)
(Females)	" X	(Fig. 12)	Positive Ego Ideal (Females)

TABLE *E*
TEST ADMINISTRATIONS

Males		Females	
Date Tested	No. in Group	Date Tested	No. in Group
Jan. 14	30	Jan. 15	30
Jan. 20	28	Jan. 21	19
Jan. 22	27	Feb. 5	6
Jan. 27	34	Apr. 6	20
		Apr. 8	15
	119		
			90

TABLE *F*
TEST DIMENSIONS

Cartoon I	Oral Eroticism
" II	Oral Sadism
" III	Anal Expulsiveness
" III	Anal Retentiveness
" IV	Oedipal Intensity
" V	Masturbation Guilt
" VI	Castration Anxiety (Males)
" VI	Penis Envy (Females)
" VII	Positive Identification
" VIII	Sibling Rivalry
" IX	Guilt Feelings
" X	Positive Ego Ideal
(Fig. 11 for Males; Fig. 12 for Females) Cartoon XI	Narcissistic Love-Object
(Fig. 12 for Males; Fig. 11 for Females) Cartoon XI	Anaclitic Love-Object
(Fig. 12 for Males; Fig. 11 for Females)	

TABLE *G*
PROFILES OF DIMENSIONAL SCORES

Male Subjects	Oral Eroticism	Oral Sadism	Anal Expuls.	Anal Retent.	Oedipal Intensity	Masturb. Guilt	Castration Anxiety	Positive Ident.	Sibling Rivalry	Guilt Feelings	Positive Ego Ideal	Narcissistic Love-Object	Anaclitic Love-Object
IM-14	++	++	++	0	+	+	+	0	++	+	0	+	0
IM-15	0	0	0	+	0	+	0	+	0	0	0	0	0
IM-16	+	0	++	0	++	0	+	0	0	0	+	0	++

TABLE *H*
DIMENSIONAL SCORING ILLUSTRATION

	Source		Oral Erotic Involvement
Subject *A*:	Cartoon I	Spontaneous Story	Strong
	Cartoon I	Inquiry	Not Strong
	Cartoon I	Preference	Not Strong
	Related Comments on other cartoons		Strong

TABLE *I*
SPONTANEOUS STORY SCORING AGREEMENT

Dimension	Obtained per cent scoring agreement	Per cent expected by chance
Oral Eroticism	100	56
Oral Sadism	96	61
Anal Expulsiveness	84	43**
Anal Retentiveness	96	44**
Oedipal Intensity	96	53
Masturbation Guilt	84	50
Castration Anxiety (Males)	100	56
Penis Envy (Females)	76	53
Positive Identification	100	79
Sibling Rivalry	92	56
Guilt Feelings	80	51
Positive Ego Ideal	100	92
Narcissistic Love-Object	92	54
Anaclitic Love-Object	100	85

	Obtained	Expected by chance
Mean	92.6%	59.5%
Median	96.0%	55.0%
Range	76%-100%	43%-92%

TABLE *J*

NUMBER OF SIGNIFICANT SEX DIFFERENCES OBTAINED VS. EXPECTED

(Total number χ^2 possible = 121)

Level of significance	Number obtained	Number expected by chance alone
5% or less	23.0	6.1
2% or less	21.0	2.4
1% or less	15.0	1.2

TABLE *K*

NUMBER OF SIGNIFICANT DIMENSIONAL INTERCORRELATIONS OBTAINED VS. EXPECTED

(Total number of r_d's possible = 225)

Level of significance	Number obtained	Number expected by chance alone
5% or less	28.0	11.3
2% or less	18.0	4.5
1% or less	13.0	2.3

REFERENCES

1. FENICHEL, O. *The Psychoanalytic Theory of Neurosis.* New York: Norton, 1945.
2. FREUD, S. Collected Papers. (Originally published 1893-1914. Trans. under supervision J. Riviere.) New York: International Psycho-Analytical Press, 1924. (Vol. I.)
3. ———. Collected Papers. (Originally published 1906-1924. Trans. under supervision J. Riviere.) London: Hogarth Press, 1924. (Vol. II.)
4. ———. Collected Papers. (Originally published 1905-1918. Trans. by Alix and James Strachey.) London: Hogarth Press, 1925. (Vol. III.)
5. ———. Collected Papers. (Originally published 1908-1923. Trans. under supervision of J. Riviere.) London: Hogarth Press, 1925. (Vol. IV.)
6. ———. New Introductory Lectures on Psychoanalysis. (Trans. by W. J. H. Sprott.) New York: Norton, 1933.
7. ———. Three Contributions to the Theory of Sex. In *The Basic Writings of Sigmund Freud.* (Trans. and Ed., A. A. Brill. Original publication 1905.) New York: Random House, 1938.
8. ———. Wit and its Relation to the Unconscious. In *The Basic Writings of Sigmund Freud.* (Trans. and Ed., A. A. Brill. Orlginal publication 1916.) New York: Random House, 1938.
9. ———. A General Introduction to Psychoanalysis. (Trans. J. Riviere.) Garden City, N. Y.: Garden City Publishing Co., 1943.
10. HAYES, S. P., JR. *Diagrams for computing tetrachoric correlation coefficients* from percentage differences *Psychometrika,* 1946, 3, 163-172.
11. MCKINLEY, J. C., (*Ed.*) An Outline of Neuropsychiatry. Dubuque, Iowa: Brown, 1944.
12. SEARS, R. R. Survey of Objective Studies of Psychoanalytic Concepts. New York: Social Science Research Council, 1943. (Bulletin 51.)

Published as a separate and in *Genetic Psychology Monographs*, 1949, **39**, 101-148.

THE ASSESSMENT OF PARENTAL ATTITUDES IN RELATION TO CHILD ADJUSTMENT* [1]

Department of Psychology, University of Southern California

EDWARD JOSEPH SHOBEN, JR.

Department of Psychology, State University of Iowa, Iowa City, Iowa

I. The problem and its background 103

II. Previous attempts to measure parental attitudes 107

III. The University of Southern California Parent Attitude Survey: *Construction of items and description of sample* 117

IV. The University of Southern California Parent Attitude Survey: The reliability and validity of the tentative scale 127

V. Summary, conclusions, and suggestions for further research . . . 137

Appendix: The University of Southern California Parent Attitude Survey 139

References 147

*Received in the Editorial Office on November 15, 1948.

[1] An abridgment and a revision of the thesis presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology of the University of Southern California, 1947.

I. THE PROBLEM AND ITS BACKGROUND

Modern dynamic psychology lays heavy emphasis on the socialization process, the process by which an individual grows from a dependent infant into an independent and dependable adult (23). One of the fundamental agencies of socialization is the family. As Teagarden (31) puts it, ". . . there is accumulating evidence that all manner of behavior deviations can be and often are accounted for by the subtleties of home relationships." It is with one aspect of this dynamic psychology of home relationships that this investigation is concerned.

The problem involves the testing of two hypotheses borrowed in part from the work of Champney (7). First, it is hypothesized that a given parent behaves toward a given child with sufficient consistency from situation to situation to differentiate himself or herself measurably from other parents. Second, it is further hypothesized that the type of characteristic parental behavior displayed is significantly related to the adjustment of the child. If these two ideas are sound, it should be possible to construct a measuring device that will differentiate the parents of maladjusted children from the parents of adequately adjusted children. In broad outline, this is the scope of the present study.

The justification for work on such a problem lies in two areas of psychological activity: (*a*) clinical work with problem children and/or their parents, and (*b*) research work on the whole matter of parent-child relationtionships. In the clinical treatment of children, the clinician must very often determine: (*a*) the extent to which parents' attitudes—and their associated behavior—have contributed to the children's difficulties, (*b*) the characteristics of those attitudes in terms of intensity and content, and (*c*) the advisability of working more intensively with the parent or with the child if time precludes doing both effectively. It is hoped that an instrument of the kind here proposed would be useful in the answering of these questions both from the standpoint of economy of time and that of comprehensiveness. Specific suggestions for further research will be made in the final chapter, but it may be pointed out at this juncture that an adequate survey of parental attitudes would be of obvious value in work on such crucial problems as these: (*a*) the effects on children of differing sets of attitudes on the parts of the father and the mother; (*b*) the relationship between attitudes toward children and the satisfactoriness of the marriage relationship for husband and wife; (*c*) the amenability of certain attitudes toward children to re-education through counseling, parent-education courses, mental hy-

giene programs, etc.; and (*d*) the relationships between certain parental attitudes toward children and such factors as socio-economic status, intelligence, degree of happiness during parents' childhood, education, and many others.

Symptomatic of the widespread current interest in the effects of parental attitudes on child development is the welter of recent publications intended for parents and educators, the dominant motif of which is the vital importance of the home for the child's psychological health. There seems to be enormous agreement among authorities that, owing to its primacy and the pervasiveness of its contacts with the child, the family exerts a tremendous influence. The statements of Faegre and Anderson (9) are typical:

> Granting the extent to which the responsibility for some types of training has been shifted to the schools, the home still offers the earliest, and in many respects the most thorough, education which a child receives. . . . We have seen that the personality of the child is emerging among all the influences of the early environment and is being shaped by them, and that the family represents the world of the child in which, long before he reaches school age, he has been meeting situations and developing ways of reacting to them. Because he is more frequently and more profoundly moved or stimulated by persons than by the inanimate parts of his environment, the home with its close association with a number of personalities, becomes the field in which the child tests out and comes to appreciate the value of certain types of behavior.

This point of view, of course, is not far from that of the psychoanalysts, who have always been insistent upon attributing the personality characteristics of the individual to the effects of his early experiences. In connection with his explanations of the development of the child's affectional life within the family, Freud has made many statements concerning the significance of parental attitudes and behavior for the child's mental health. He insists, for example, that it is the "neuropathic parents, who usually display a boundless tenderness, who often with their caressing awaken in the child a disposition for neurotic diseases" (12). In like manner, Flügel (10) asserts that the child whose parents are either too severe or too careful in bringing him up is likely to be rebellious toward not only his parents, but toward all adult authority.

In keeping with such assertions as these is the recent spate of popular literature on "momism," the burden of which is that modern youth, especially modern male youth, has been seriously damaged psychologically by maternal dominance on the one hand and maternal over-protectiveness on the other.

Philip Wylie's (36) has been the most vitriolic attack on the "moms" of the land, but his strident voice has received sizeable support from E. A. Strecker (29), the prominent psychiatrist, who has based his indictment on his psychiatric experiences in dealing with men who suffered psychological breakdowns during World War II. His thesis is that many men have proved low in resistance to psychological stress owing to the influence of their "moms," who have "failed in their elementary mother function of weaning their offspring emotionally as well as physically." While he hastens to differentiate between "moms" and "mothers" on the basis of the adequacy with which psychological weaning has been accomplished, Strecker makes it plain that "momism," in his estimation, is a widespread blot on the nation's psychological landscape.

Such points of view are essentially verified by clinical and experimental investigations of parent-child relationships. Witmer (35), for example, found that parent attitudes are an important prognostic index in clinical work with children and that clinical success is to a large degree dependent upon a favorable psychological climate for the child in his home. Symonds (30) reports suggestive correlations between the characteristic behavioral trends of children and parental tendencies to reject the child or to overprotect him. Hattwick and Stowell (16) discovered that children's work habits in school are related to such home influences as "babying," pressure and insistence on high standards, and "good" adjustment among the various members of the family constellation.

The classic work by Levy (20) on maternal over-protection indicates definite relationships between patterns of child adjustment and certain home variables such as excessive physical contact, prolongation of infantile care, prevention of independent behavior, maternal dominance, and maternal indulgence. Levy shows, moreover, that overprotection as a parental attitude is in large part determined by the particular psychological experiences of the mother herself in the course of her contacts with her own family.

In like vein, Karlin (17) reports that 67 per cent of the mothers of a group of neurotic children were themselves diagnosed neurotics. Foster (11), Sewall (27), and Ross (26) in their studies of sibling jealousy found strong indications that jealousy of sufficient magnitude to amount to a serious mental hygiene problem is a function of certain parental behavior such as nagging, inconsistent discipline, etc.

From these clinical studies emerges decided agreement on the following points: (*a*) Childhood personality and behavior problems seem to be related to parental policies and their manner of execution. (*b*) Over-protection, re-

jection, repressiveness, severity, domination, and undue submissiveness seem to be the parental traits which are associated with children's difficulties. (c) The provision of a home in which the child can grow up feeling emotionally secure seems to be the basic requisite in the socialization of the child.

Nor are these generalizations contravened by experimental studies. Hattwick (15) reports impressive correlations between the behavior of preschool children as determined by teachers' ratings and parental attitudes as measured on the basis of home visits. Grant, in an unpublished M.A. thesis cited by Updegraff (32), reports a similar study in which definite relationships between the behavior of preschool youngsters and patterns of parent behavior were found. The later and more comprehensive studies of Radke (25) and Lafore (18) quite definitely confirm the conclusions of Hattwick and Grant: The behavior and attitudes of the child are in large degree determined by the behavior and attitudes displayed toward him by his parents. Experimental studies by Anderson (1) with junior high school children and by Meyers (22) with high school students resulted in the same general finding and indicate that the importance of parental attitudes in shaping the young personality is not restricted only to the younger child. Likewise, Carpenter and Eisenberg (5) and Goodwin Watson (33) found that there are highly suggestive relationships between adult personality patterns and histories of exposure to various kinds of parental attitude.

From the general literature, then, and from clinical and experimental studies it seems safe to conclude that the family as an agency of socialization has an influence on the child that reverberates throughout his life span, sharing importantly in the determination of his personality and the behavior reflected from it. Out of this general notion can be developed the two hypotheses mentioned at the beginning of this section: (*a*) that a given parent behaves toward a given child from situation to situation with sufficient consistency to differentiate himself measurably from other parents, and (*b*) that such characteristic parental behavior is significantly related to the success or failure of child adjustment.

To test these hypotheses, an inventory-type "test" of parent attitudes, the University of Southern California Parent Attitude Survey, was constructed against the criterion of child adjustment. As a background for the discussion of the procedures utilized, previous attempts to measure parent attitudes systematically will be reviewed in the next section.

II. PREVIOUS ATTEMPTS TO MEASURE PARENTAL ATTITUDES

In spite of the importance of the problem of parent-child relationships as factors in the development of personality, in spite of the wealth of study that has been reported in the literature, and in spite of the prominence of parental attitudes and behavior as background for the aberrant behavior of problem children seen in clinics, there has been developed to date no adequate measure of parental attitudes that would be of value for clinical purposes or that would aid in more precise inquiries into the dynamic patterns of parent-child relationships. Moreover, there has been surprisingly little work done toward achieving this end. Actually, there are in the literature only six studies devoted to the development of four measuring devices that merit attention. These involve three different approaches to the assessment of parent attitudes: (*a*) direct observation of behavior, (*b*) rating scales, and (*c*) inventory-type questionnaires. The significant work related to each of these approaches may be briefly considered here.

A. Measures Derived from Direct Observation

1. *Merrill's Study*

Merrill (21) attempted to study and measure the stimulus properties of maternal behavior toward preschool children in standardized play situations. Viewing the behavioral interaction of mother and child through a one-way screen, this investigator kept a running record of her observations by notations taken every five seconds during two 30-minute sessions, utilizing 32 predetermined behavior categories, 11 of which were finally analyzed, for her recording system. The 11 variables studied were lack of contact, structurizing, structurizing a change in activity, teaching, interactive play, helping, directing, interfering, criticizing, coöperation, and noncoöperation.

Her subjects were 30 mothers divided equally into a control and an experimental group, matched in terms of the predominate behavior shown in the first of the play sessions between mother and child. The control mothers were tested in two half-hour play periods under identical conditions; the experimental mothers were told before the second period that their children's play in the previous session had not shown full realization of his capabilities. This was presumed to increase the mother's motivation to induce her youngster to play "well" and thus to introduce an element of pressure into the total situation.

The control mothers showed consistent trends in behavior from the first

to the second session. The experimental group, however, showed significant increases in the direction of more directing, interfering, criticizing, and structurizing changes in activity.

Three specific aspects of the problem were considered in Merrill's analysis of her data: (*a*) the consistency of maternal behavior, (*b*) the effects of increased motivation on maternal behavior, and (*c*) individual differences in maternal behavior.

With respect to consistency, Merrill compared the records for the control mothers' first and second play sessions. Since the conditions for this group were identical for the two periods, it was hypothesized that no differences should appear. The hypothesis was borne out by the fact that no significant differences in averages or variabilities occurred for any of the 11 categories with one exception; the variability of the *coöperating* variable showed an increase significant at the 1 per cent level. An examination of individual records showed, however, that three cases of comparatively large change accounted for this difference. Rank order correlations between scores in the first session and scores in the second also indicated a high degree of consistency in maternal behavior.

Merrill concludes, therefore, with respect to the consistency issue, that it ". . . is evident . . . that the experimental conditions offered the mothers an opportunity to display a consistent form of behavior" (p. 43). She does not, however, attempt to establish a relationship between home conditions and the experimental conditions and to predict home behavior from the experimental findings. Since every effort was made to make the experimental situation analogous to a home situation, it seems logical to assume that her caution on this point is more a matter of good scientific procedure than any conviction that the differences between home and experiment were so large as to invalidate the study for any conditions other than those obtaining in the experiment.

In studying the effects of increased motivational pressures on maternal behavior, Merrill told the members of her experimental group when they returned for their second session that the previous play period had not shown their children performing up to the level of which they were capable. This was calculated to increase the mother's motivation to have her child play "well" during the interview.

The categories which clearly showed differences from the first to the second session under these conditions were *directing* and *interfering*, both of which were significant at the 5 per cent level of confidence. Two other variables, *suggesting* and *criticizing*, were significant well beyond the 20 per cent

level, approaching the 5 per cent level of significance. These differences were between means on the basis of the *t*-test. When variabilities were compared from session to session in the experimental group, all the 11 categories showed increase significant at the 5 per cent level and beyond with the difference in *criticizing* significant at the 1 per cent level. Inspection of the individual records indicated that only two of the 15 mothers in the experimental group had failed to respond to the pressure in some detectable way. *Merrill's conclusion from these findings, apropos of the problem of the* effects of motivation on maternal behavior, are stated in this wise:

> *In general, a mother's relationship with her child appears to be* influenced and changed by her motivation to have that child appear to the world in the best possible light. When this motivation is restricted to a specific situation, delimited in time, and defined in terms of performance desired, the mother tends to assume direct control of the child's actions, and to impose her own standards, rather than to interact *with him in such a way as to advance his ability to think and act independently and to foster his autonomy within the limits of necessary* adult guidance (p. 46).

Examination of the individual records in an effort to investigate the individual differences shown among the 30 mothers revealed a wide range of behavioral patterns, no two of which were alike. Merrill reports three cases as examples, quoting from the shorthand transcription of the experimental sessions as well as from the scores attained in terms of the frequency of occurrence of the several variables of the study. A brief characterization of each of the cases will indicate the range of individual differences found.

One mother in the experimental group attempted to exhibit her child's abilities even in the first session by having him spell his name, recall telephone numbers, and build various complex structures with building blocks. The relationship between mother and child was extremely close and the youngster, a four-year-old whose "speech and manner were that of a miniature adult," was obviously tremendously dependent on his maternal parent. When subjected to the motivational pressure of the second session, this mother was "characterized by an almost desperate attempt to force the child to display his capabilities."

A second mother in the experimental group displayed behavior almost the exact reverse of the above. This woman showed what seemed to be a genuine interest in and enjoyment of her child as an individual in his right. In the first session her interaction with the youngster was based on a full respect for his autonomous individuality and at the same time on a recognition

of his need for guidance. She scored quite high on the *teaching* variable, but her instruction was related to stated or sensed desires on the part of the child for information. In the second session, her frequency of *teaching* tended to go down and her frequency of *directing* and *interfering* to go up, but the changes were not large and there were no noticeable changes in her behavior as it was grossly observed by the experimenter. The child obviously looked upon his mother as a splendid companion, showed an easy adaptation to the play situation, and carried on his play on a highly constructive level. He revealed no particular dependence on his mother but at the same time was quite considerate of her as evidenced by a spontaneous apology for spilling water on her shoes at one point in the play.

The third mother, also a member of the experimental group, was a nervous, high-strung person, compulsively maintaining constant contact with the child. Her behavior in the play situation alternated between the very friendly and permissive and the very domineering and irritated. The latter tended to show themselves whenever the child deviated in the slightest from the mother's standards for him. She was uncertain and inconsistent in her interaction with the youngster. The child, a three-year-old, played quite adequately but within a restricted range of interests, ignored his mother or showed indifference to her, and revealed a firm determination to follow his own inclinations in play in spite of any stimuli she might present.

2. *Evaluation of Merrill's Study*

This study has been reviewed in some detail since it is the only one of its kind to be found in the literature and because of the light it throws on parent-child relationships both methodologically and in terms of findings. Certainly, from the point of view of describing maternal behavior, this represents a fruitful approach, combining some of the concepts of the clinic with the precision of the laboratory.

Time and cumbersomeness, however, make the technique difficult to adapt for clinical use—a criticism which is, of course, quite irrelevant to the purpose for which the study was developed—and it is essentially descriptive rather than dynamic. That is, Merrill's method of direct observation permits the construction of categories according to which maternal behavior may be classified; it does not yield, except possibly by brain-wrenching inferences, much information on the affective tone of the parent-child relationship, nor does it directly reveal much about the attitudes of the parent. Actually, the method seems to be essentially a surrogate for home visitation and requires the same kind of validation as any other kind

of observational assessment before it can be considered acceptable as an instrument by which light can be thrown on the socialization of the child through the agency of its family.

B. Rating-Scale Measurements of Parent Behavior

By far the most definitive attempt to date to measure parent-child relationships is that of Horace Champney (6) and his co-workers at the Fels Foundation to construct rating scales to be used by home visitors in assessing parent behavior in the home. The Fels Parent-Behavior Rating Scales provide for the evaluation of parent behavior in terms of 30 variables defined as continua characterized by concretely expressed cue-points which regulate the kinds of ratings assigned by the visitor to the parent. The 30 variables follow:

1. Adjustment of home: Maladjusted—Well adjusted.
2. Activeness of Home: Inactive—Active.
3. Discord in Home: Harmony—Conflict.
4. Sociability of family: Reclusive—Expansive.
5. Coördination of household: Chaotic—Coördinated.
6. Child-centeredness of home: Child-subordinated—Child-centered.
7. Duration of Contact with Parent: Brief contact—Extensive contact.
8. Intensity of Contact: Inert—Vigorous.
9. Restrictiveness of regulations: Freedom—Restriction.
10. Readiness of enforcement: Lax—Vigilant.
11. Severity of actual penalties: Mild—Severe.
12. Justification of policy: Arbitrary—Rational.
13. Democracy of policy: Dictatorial—Democratic.
14. Clarity of policy: Vague—Clear.
15. Effectiveness of policy: Unsuccessful—Successful.
16. Disciplinary friction: Concordant—Contentious.
17. Readiness of suggestion: Non-suggesting—Suggesting.
18. Coerciveness of suggestion: Suggestions optional—Suggestions mandatory.
19. Accelerational attempt (pushing): Retardatory—Acceleratory.
20. Babying: Withholds help—Over-helps.
21. Protectiveness: Exposing—Sheltering.
22. Readiness of criticism: Uncritical—Critical.
23. Direction of criticism: Disapproval—Approval.
24. Readiness of explanation: Thwarts curiosity—Satisfies curiosity.
25. Solicitousness for child's welfare: Nonchalant—Anxious.
26. Acceptance of child: Rejection—Devotion.
27. Understanding of child's problems: Obtuse—Keen.
28. Emotionality toward child: Objective—Emotional.
29. Affectionateness toward child: Hostile—Affectionate.
30. Rapport with child: Isolation—Close rapport.

The questions that immediately arise are those related to the reliability and validity of the scales and to the amount of overlap among them, i.e., the degree of empirically determined intercorrelation among the 30 scales. The most relevant data bearing on these matters are presented in a study by Baldwin, Kalhorn, and Breese (2), which used the Fels scales as its basic tool.

These researchers report data gathered from six visits, one every six months for a period of three years, made by a single home visitor to 125 homes. The reliability of the Fels scales is inferred from the intercorrelations among the ratings made successively by the single rater. The obtained coefficients range in magnitude from .62 to .90. While this is not spectacular, it is sufficient for most uses to which the scales might be put.

Champney (6) has elsewhere reported essentially confirmatory data on the reliability of the scales. Correlating a home visitor's original ratings with re-ratings made after a two-day interval, he obtained a root-mean-square value of .85 for all 30 variables. When ratings between two home visitors were correlated, the root-mean-squares ranged from .14 to .72 with an average of .49 raised by the Spearman-Brown prophecy formula to .66. A conclusion of reliability adequate to most practical and "pure" research work seems justified.

The crucial issue of validity is more difficult to dispose of. Baldwin, Kalhorn, and Breese attempt to demonstrate the validity of the scales by showing that: (*a*) the variables are reasonably interrelated; (*b*) the parents studied could be separated into groups which showed differential behavioral uniformities; and (*c*) the children of people in any one group of parents showed more behavioral uniformity than was the case in the total population studied.

The data presented by Baldwin, Kalhorn, and Breese relating to the interrelationships of variables as an indication of validity also answer the question relating to the amount of empirically determined intercorrelation of variables.[2] By a technique of "syndrome analysis," these investigators selected from a table of intercorrelations all sets of variables whose minimum intercorrelation was .60 or higher. Syndromes so formed were then intercorrelated and combined if the r between them was greater than .40. By this process, three central syndromes of parent behavior emerged which were labeled *Democracy in the home, Acceptance of child, and Indulgence.* These

[2] From their matrix of intercorrelations, it is obvious that some of the scales are very highly interrelated (e.g., *Justification of policy* and *Democracy of policy*, .92); whereas others bear very little relationship to each other (e.g., *Restrictiveness of regulations* and *Activeness of home*, .00).

syndromes, looked upon as patterns of characteristics around which parent behavior may be organized and which reflect emotionalized attitudes, philosophies of child management, and/or parental personality traits, may be looked upon as internal evidence for the validity of the scales. In other words, the scales afford some kind of statistical patterning of parental behavior into reasonably independent clusters. Unfortunately, the intercorrelations of the syndromes are not given, but since the method of deriving the syndromes involved the combining of those variables which showed high degrees of relationship, it seems likely that the syndromes themselves are related only in terms of relatively low values of r because the variables composing one syndrome have only low relationships with those composing any other syndrome. This means that there is a high probability of factor validity.

In terms of the uniformity of parental behavior which constitutes the second line of evidence for validity, Baldwin and his co-workers analyzed a number of intensively studied homes as typical of the behavior corresponding to ratings on the scales. The data do not lend themselves to correlational statement, but it is apparent that the syndromes, statistically derived on the basis of internal relationships, are related to objective behavior that justifies the naming of the syndromes and argues for the usefulness of the scales in the prediction of patterns of actual parent behavior.

The evidence for uniformity of child behavior within the groups classified according to the basic types and syndromes as indicated by the ratings rests on the extensive testing program and opportunity for observation of children which are part of the Fels Institute's longitudinal child study project. Analyzing the intellectual, social, and emotional development of the 125 children in terms of psychometric examinations at six-month intervals and yearly month-long periods of observation in the Fels nursery school, Baldwin, Kalhorn, and Bresse found definite trends characterizing and differentiating the children whose parents fell within the various categories of parental behavior. As they put it:

> . . . We have found that the children, selected on the basis of parental behavior, do show consistent uniformities. For some of the groups, the *democratic,* the *indulgent,* and the *actively rejected,* the patterns are clear cut; for the others, they are only suggestive. The existence of these patterns lends further support to the syndrome analysis of parent behavior and leads us to believe that the method is empirically fruitful (2, p. 70).

Since the syndrome analysis rests upon the rating scales, it seems legitimate to infer that evidence for the validity of the syndromes derived from the

scales is evidence for the validity of the scales themselves. Again it is highly unfortunate that the data upon which the argument for validity is based are not expressible in strict quantitative terms. Nonetheless, the data are no less convincing for their qualitative presentation, especially since they rest on the firm foundations of test scores and nursery school ratings by trained observers.

The Fels Parent-Behavior Rating Scales, then, seem to be effective and adequate measuring devices for the study of parental behavior. Have they any drawbacks that would warrant work on another type of instrument rather than on the refining and sharpening of this one?

The answer to this question is to be found again in the spheres of utility and economy. From the clinical point of view, such ratings are of inestimable value, but practically they are capable of making only a negligible contribution. To make the ratings requires a half-day visit by a trained home visitor. Both in terms of time and personnel, such a practice is not feasible in the great bulk of clinical work that is practiced currently. Secondly, one wonders if the presence of a visitor with a clinic affiliation in a home from which the clinic was dealing with a maladjusted child might not create a situation badly contaminated with a new variable that might well express itself in any direction. In homes taking voluntary part in a longitudinal research program on child development in which the flavor of the clinic is conspicuous by its absence, the presence of the home visitor probably affects the behavior of both parent and child little more than does that of any other guest. Here the validity of the scales is demonstrable and demonstrated. It may be a problem for empirical research to determine whether or not the same satisfying evidence for validity is derivable from the same method applied to a clinical situation, but it does not appear to be a problem on which to lavish one's hope if one's interest is in the development of a clinically useful instrument for the measurement of parent behavior.

C. Questionnaire Approaches to the Measurement of Parent Behavior

Curiously enough, in the light of the numerous self-inventory tests of attitudes, interests and personality, there have been almost no attempts to attack the problem of measuring parental patterns by this means. True, there have been a number of 10-or-20-item questionnaires published in such popular periodicals as *Parent's Magazine* and the *American Magazine*, purporting to indicate whether the individual taking the "tests" is a "good" or "poor" parent. These can be dismissed without comment beyond the

mention of the lack of standardization, the shortness of the tests, the baldness of the items, permitting sophistication of either a deliberate or an unconscious kind, and the arbitrariness of both the categories dealt with and the responses that are supposedly "correct."

So far as the literature reveals, there have been only two gestures in the direction of measuring parental attitudes by self-inventory methods that are worthy of discussion.

The first of the two was the study by Laws (19) in 1925, in which a four-part questionnaire was devised and administered to 50 mothers belonging to a child-study group. No data on reliability or validity are published with respect to this early inventory. So far as can be determined, no effort was ever made to set up norms in a systematic manner or to attempt a differentiation of groups with the instrument. By the standards of 20 years later, it tends to be a sorry piece of work, but it must be given credit for pioneering into a wilderness which has not yet been cleared of obscuring foliage and impeding underbrush.

The other scale is that developed by Stogdill (28) in connection with his study of differential attitudes among groups of adults toward various aspects of child behavior. This inventory consists of 60 items to be rated from one to 10 according to how seriously or unfavorably the adult rater feels the behavior described in the item affects a child. Sample items include such things as:

> Scolding the child for asking questions about sex.
> Keeping the child most of the time with adults.
> Telling the child one's troubles and worries.

By contrasting the responses of parents with those of mental hygienists, Stogdill hoped to find areas of significance in adult-child relationships and to diagnose parental attitudes that varied significantly from mental hygiene principles. An error in his method of presenting the test to the two groups has been pointed out by Goodwin Watson (34) which invalidates his somewhat skimpy findings completely. The test was never standardized in the sense of being given to a representative sample of the population so that norms could be established and the reliability of the instrument determined to say nothing of its never being administered under conditions that would permit an estimate of its validity in terms of some kind of external criterion other than the responses of mental hygienists.

D. Conclusions from the Previous Attempts at Measurement

A review of the previous attempts at the measurement of parent attitudes and behavior shows that there has been a paucity of effort expended in this direction with the exception of two studies: Merrill's measurement of mother-child interaction by means of direct observation methods, and the Fels Research Institute's work on rating scales to be used by home visitors.

Both of these approaches have great value methodologically and for research purposes of a variety of kinds. Neither, however, has been set up within the context of a study specifically designed to test the hypothesis that child adjustment is in large part a function of parental attitudes; and neither contributes much toward the goal of an instrument which would permit an assessment of parental factors as they operate in homes from which problem children come to the attention of clinicians. Such an instrument would have to be economical in terms of the amount of time required for its use and of sufficient reliability and validity to allow clinical workers to make reasonably accurate inferences as to the extent and dynamic nature of home influences when they are of importance in the etiology of the child's maladjustment.

Since there is a vast number of self-inventory tests of attitudes, interests, and personality in common use, it is a little startling to find that there is no worthwhile questionnaire available for the assessment of parental attitudes.

III. THE U. S. C. PARENT ATTITUDE SURVEY: CONSTRUCTION OF ITEMS AND DESCRIPTION OF SAMPLE

This inquiry, it will be recalled, has two underlying purposes: The first is to test the twin hypotheses (*a*) that a given parent behaves toward a given child with sufficient consistency from situation to situation to differentiate himself measurably from other parents, and (*b*) that the success or failure of the child's adjustment is in large part a function of the parental behavior to which he has been exposed. The second is to construct an instrument that will be useful in the assessment of parent attitudes both for clinicians and for research workers dealing with the problem of parent-child relationships in the socialization process. These dual purposes, it must be pointed out, are by no means separate. They are closely united by virtue of the fact that the attainment of the first is prerequisite to the attainment of the second. In other words, the construction of a measure of parental attitudes against a criterion of child adjustment constitutes a test of the hypotheses initiating the study.

This section and the next will report the processes carried out in devising the University of Southern California Parent Attitude Survey, a self-inventory type scale designed to assess parent attitudes (as indicators of parent behavior) in relation to behavior and/or personality problems in children. This section will describe the construction of the items and the composition of the sample upon which the items were tested. The next section will report the results of that testing in terms of the reliability and validity of the scale.

A. Construction of Items

A self-inventory type of measuring device was decided upon as the kind of instrument to be built because of ease of administration, the objectivity with which such a scale can be validated and standardized, and the previous success of attitude scales like Thurstone's and similarly constructed personality tests like Guilford's and the Minnesota Multiphasic Personality Inventory. It was tentatively decided to call the assessment device here considered the University of Southern California Parent Attitude Survey, largely because such a title seems to sound innocuous and disguises to some extent the purpose of the scale, which is to separate the parents of problem children from those of non-problem children in those cases where parental attitudes are probably of etiological significance in a child's psychological disturbance. More specifically, it is intended that the scale provide a point

of departure for the analysis of patterns of attitude and specific home variables that are influential in shaping the behavior and personality of the child.

The first problem in the construction of the Survey was the writing of items. It was hypothesized that to be discriminatory, items must meet two criteria: (*a*) they must deal with significant aspects of the parent-child relationship, but (*b*) they must be adequately disguised in order to minimize sophistication.

On the basis of this reasoning, a number of different ways to cast items presented themselves. The most obvious method was to write statements descriptive of individual parent behavior which could be responded to by classifying the statements as True or False. For example, an item like, "I very frequently spank my child when he disobeys," could be responded to with "True," "False," or "?," and statistical tests made of the discriminatory power of the items. It was decided, however, that such a form for the presentation of items would call, in effect, for confessions by parents of various kinds of behavior to which they might be quite hesitant to admit. Not only would this encourage deliberate sophistication but it might well tend to impair rapport in a clinical situation by putting the parent on the defensive with the clinician. Consequently, this method of presenting items was dropped.

It must be pointed out, however, that this decision was made on a purely a priori basis and without experimental justification. Nonetheless, the action seems allowable on the basis of logic and clinical experience with other tests.

A second type of item that was considered was a self-rating by the parent of some complex home variable. For example, the parent could be requested to rate his or her home according to its child-centeredness on a scale ranging from "Whole household revolves around child" to "Household organized around adult needs and interests with child required to fit in as best he can." This type of item presents the rather interesting experimental problem of adapting Champney's *Fels Parent Behavior-Rating Scales* for self-administration by parents themselves. Two difficulties immediately presented themselves, however, which led to the discarding of this approach. One was the omnipresent problem of reporting errors because of sophistication, an inability to make accurate judgments through lack of insight into home conditions, or the effects of affectively toned phrasing in the description of the cue points along the scale. The second factor militating against this approach was the numerous practical difficulties in the way of establishing the reliability of such a measuring device. Asking for re-ratings by parents

presents obstacles both because of the possibility of actual change in the home and because of changes in the verbalizations of parents owing to such influences as reading prompted by the taking of the scale or talks with the administrator, to say nothing of the problem of actually getting parents to make the ratings twice. The establishment of reliability by correlating paternal and maternal responses was considered but discarded because the differential judgments concerning the home made by the father as against the mother might well be an important factor in a valid assessment of the home. Internal consistency was not considered because the variables making up the scale would not be amenable to frequent enough restatement to allow a stable estimate of reliability by this method. For these two reasons, especially the first, the idea of a self-rating was not followed up in spite of its interesting possibilities and the attractive validation procedure of comparing parent ratings with the ratings of a trained home visitor.

A third possibility was to set up hypothetical situations between child and parent and ask the mother to respond to various alternative courses of action which she might take. For example, the parent could be asked to mark the alternative which most nearly describes her probable behavior in the following situation.

> Your six-year-old boy has refused to eat his vegetables at dinner. You *most likely would*:
>
> 1. Insist that he stay at the table until he had finished them.
> 2. Remove the dishes without comment when the adults had finished their meal.
> 3. Punish him in some way.
> 4. Coax him to eat them.

This again seemed to be a worthwhile attack on the measurement problem, but the cumbersomeness in terms of sheer size of the scale, the difficulty of getting situations that would be representative for all the homes that would come within the purview of the test, and the possibility of becoming stimulus bound by the specificity of the situations finally outweighed the appeal of the method. It is suggested, however, that a study of the effectiveness of this type of approach be made at some future date.

A fourth kind of item consisted of general statements of parental policy or attitude intended to tap the affectively toned points of view that parents might have regarding children. It was hypothesized that the most probable referent for "children" for parents taking the Survey would be their own youngsters and that in consequence responses to items of this type would reveal intrafamilial attitudes of significance for child development. For

example, such general statements as the cliches, "Children should be seen and not heard," and "Children need some of the natural meanness taken out of them" might well be discriminatory on the basis of their revealing affective attitudes determining to some extent both the behavior of parents toward their children and the general psychological climate of the household. Furthermore, such items avoid the specificity of the situational approach described above, the obstacles to the determination of reliability and the accurate reporting of home variables by parents judged to be inherent in self-ratings, and the virtual invitation to sophisticate felt to be characteristic of items asking for "confessions" of parental behavior toward children. In addition, a test composed of items of the sort here proposed would be easy to administer, deal with the significant variables of the parent-child relationship, and be easily subjected to tests of the significance of differences between the responses of parents of problem children and those of parents of non-problem children.

On the other hand, this kind of item is particularly vulnerable to semantic factors. "Purr words" like "maturity" might easily elicit an affectively colored response from a parent quite at variance with the parent's actual feelings with respect to "maturity" in his or her own youngsters when designated by some other term.

It was felt, however, that this difficulty could be overcome more readily than the difficulties that have been described in connection with the other possible approaches that were considered. Consequently, this attack was the one selected. Again it must be emphasized that the type of item finally selected was the result of an armchair analysis of the various possibilities and hence subject to the errors that can occur from the use of logic without the buttressing of empirical test. On the other hand, there seems to be no evidence readily available that would indicate that this procedure is either unjustified or at odds with that employed by the builders of other similar types of measuring instruments like personality tests or attitude scales. While it seems necessary to call attention to the fact that the method of selecting the type of item to be used was chosen on the basis of a logical consideration of the different approaches possible, it does not seem necessary to couch the statement of this fact in the language of apology.

When this mode of approach to the problem of measuring parental attitudes had been decided upon, 148 items were written for inclusion in the original scale.[8] These consisted of statements of general attitudes toward

[8]The term *original scale* is here used to indicate the battery of 148 items that was composed on an armchair basis for presentation to a trial sample. It is used

children to which the parent could respond by indicating strong agreement to strong disagreement. Since it was felt that these statements could tap emotionally colored attitudes, it was hypothesized that if the parent were urged to work rapidly, the results would come closer to mirroring the actual *feelings of the parent toward his or her children.* The recommendation to work rapidly was therefore included in the instructions to respondents and the various administrators of the original scale were requested to emphasize this matter.

B. Description of the Sample

The original scale was administered to 100 white, urban mothers, 50 of whom were the parents of problem children and 50 of whom were the parents of non-problem children. The definition of "problem child" rested on any one of three criteria: the child was receiving clinical help for some personality or behavior problem; the child had come into the custody of the juvenile authorities at least twice; or the child's mother registered a complaint about the youngster clearly indicating that she would like to have clinical help with her child if it were available or if she could afford it. On the other hand, the definition of "non-problem child" rested upon the child's never having received clinical attention, never having been taken into custody by the juvenile authorities, and having the mother deny that clinical help was either necessary or desirable with her child.

In the case of the problem children, these data were easily obtained. The mothers were invited to participate in the study because of the very fact that their children were receiving clinical attention or had been apprehended by the juvenile authorities. In eight cases the children did not fall into either of these two categories, but the mothers complained about their children's personalities or behavior and specifically made inquiries about the possibility of clinical help. With the non-problem children, the mothers indicated that their youngsters had never been apprehended for delinquency, had never received psychological counseling from a psychologist, physician, or child guidance specialist, and they also indicated that they felt in no need of help from such workers in connection with their children.

In administering the original scale, all subjects were assured of as much

---

in contrast to the term *tentative scale* which refers to the items which were selected from the original 148 on the basis of the significance of differences between the responses of the parents of problem children and those of non-problem children in the sample studied. *Tentative scale* rather than *final scale* is used because it is felt that the test needs further work with a much larger sample before it can be looked upon as in final form.

anonymity as possible. In the case of the mothers of non-problem children this amounted to complete protection from having their identity revealed. In the case of the mothers of problem children, they were assured that no one besides the clinician whom their children were seeing would have any idea of who had filled out the form. In no case was it necessary for the mother to sign her name, but such a course was left open. Nine mothers of non-problem children and four mothers of problem youngsters indicated their names and made requests for information about the results of the testing.

Subjects were obtained from the following situations: The mothers of non-problem children (hereafter referred to as the non-problem group) were obtained from: (*a*) the out-patient pediatric clinic of a Los Angeles hospital, where 15 mothers brought children for fractures, dental caries, hypertrophied tonsils, or colds, and where questioning indicated that the criteria for definition as non-problem children were met; (*b*) a meeting of the Parent-Teacher Association of a junior high school in a Los Angeles suburb, where 26 mothers filled out the Survey and indicated that their children were of the non-problem variety as defined in this study;[4] and (*c*) nine friends and acquaintances of the writer whom he asked to fill out the Survey and return them by mail in envelopes without return addresses in order to insure anonymity in these cases as in all others.

The problem group of mothers came from the following sources: juvenile section of the county probation department, 20; a municipal child guidance clinic, 11; University of Southern California Psychoeducational Clinic, 6; an out-patient psychiatric clinic for children, 6; and the private practice of two child guidance experts, 7.

The two groups of mothers were compared for the following factors: age, education, socio-economic factors, marital stability, and incidence of psychological or psychiatric treatment for parents, i.e., the mother herself and her husband. The data on these factors will be reviewed at this point.

1. *Age*

The mean age of the mothers in the problem group is 37.30 years as contrasted with a mean of 33.76 years in the non-problem group. Although this difference yields a t of 2.68, significant beyond the 1 per cent level, it seems doubtful if this difference of four years in the middle thirties is very mean-

[4] Thirty-two mothers were present at the meeting and filled out the Survey, but six of the cases were not used because some doubt could be thrown on the classification of the children as definitely problem or non-problem according to the criteria used in the present study.

ingful psychologically. Since it might mean, however, that the children of the non-problem mothers were a good bit younger than those of the problem group, which in turn could mean that children classified as non-problem fell into that group only because they had not yet lived long enough to show behavior similar to that of the children classified as problem, the mean ages of the problem and non-problem children were computed. These turned out to be 10.54 and 8.98 respectively. The *t*-test yielded a value of .80, which is insignificant.[5] The differences, then, in the ages of the children and consequently in their opportunity to develop behavior or personality problems are inconsequential for the purposes of the present study. The age of the mothers, however, is a poorly controlled variable, although its psychological significance, if any, is somewhat obscure.

2. *Education*

The educational attainments of the mothers are shown in Table 1. As can be seen, the *t* value here is 3.08, which is significant well beyond the 1 per cent level. This constitutes a serious departure from the ideal of perfectly matched groups since Baldwin, Kalhorn, and Breese (2) have shown that education is highly related to democratic forms of child management. This deficiency in the sample is further underscored when the education of the husbands is analyzed. These data are presented in Table 2. Here the difference is significant beyond the 5 per cent level. Taken together

TABLE 1
EDUCATION OF MOTHERS

Problem			Non-Problem			
N	*M*	*SD*	*N*	*M*	*SD*	*t*-value
50	10.86	2.88	50	12.34	2.79	3.08

TABLE 2
EDUCATION OF HUSBANDS

Problem			Non-Problem			
N	*M*	*SD*	*N*	*M*	*SD*	*t*-value
42	10.86	3.48	49	12.45	3.47	2.16

with the very significant difference between the two groups of mothers, these data indicate that the educational backgrounds in the homes of the problem children and those of the non-problem children are quite different, a fact

[5]The sigma values for these groups are as follows: problem mothers, 7.29; non-problem mothers, 5.76; problem children, 2.41; non-problem children, 2.69.

which militates against the purity of the study. Some of the attitudinal differences showing themselves in the Survey may well be a function of education rather than of "good" or "poor" parent behavior, thus lowering the discriminating power of the instrument when education is held constant. This point must be kept in view when conclusions are drawn about the usefulness of the Survey.

3. *Socio-Economic Level*

Socio-economic factors in the two groups were evaluated in terms of two criteria, income level and occupational level. Table 3 shows the distribution of the two groups according to the way in which the mothers checked a list of annual incomes in terms of the closest approximation to their own. By the chi-square test, the probability is less than 10 per cent that the differences shown could have arisen from other than chance. In addition, Table 3 shows that a rather sizeable spread of incomes is represented in the sample.

TABLE 3
INCOME LEVELS FOR PROBLEM AND NON-PROBLEM GROUPS

Annual income	Problem group	Non-problem group
$1,500.00	6	4
2,000.00	5	6
2,500.00	4	7
3,000.00	5	6
3,600.00	8	9
4,800.00	11	10
6,000	7	6
Not Reported	4	2
Totals	50	50
	$\chi^2 = 2.57$	$P = > .90$

The sample, however, *does not conform to the distribution of national incomes*, which follows a strongly positively skewed curve.

In Table 4 the occupational levels of the two groups are shown according to the Goodenough Occupational Classification (13). Here again the chi-square test shows that the differences are not significant. The odds are almost four to one, however, that some factor other than chance accounts for them. Since the level of significance is quite low, Table 4 reflects adequate matching in the two groups, but greater similarity between the groups would be decidedly desirable.

While it is rather curious in the light of the highly significant differences between the problem and non-problem groups and their husbands in terms of education, it can be fairly concluded that the differences are of little

TABLE 4
OCCUPATIONAL LEVELS FOR PROBLEM AND NON-PROBLEM GROUPS

Occupational levels		Problem group	Non-Problem group
I	Professional	4	7
II	Semi-professional and managerial	8	10
III	Clerical and skilled trades	15	13
IV	Farmers		1
V	Semi-skilled occupations	11	11
VI	Slightly skilled occupations	3	4
VII	Day Laborers	3	4
VIII	Unreported	6	0
	Totals	50	50
		$\chi^2 = 9.57$	$P = > .20$

consequence with respect to socio-economic level as indicated either by income or by occupational classification.

4. *Marital Stability*

Using divorce as an indicator of marital stability, the problem group shows an incidence of 16 divorces with four of those 16 cases having been divorced twice. In the non-problem group only eight divorces are reported with only one case having been divorced twice. The evaluation of these facts presents difficulties. A number of studies (3, 4) have stressed the importance of broken homes and family discord in childhood maladjustment, and from the common principles of mental hygiene one would predict that an individual undergoing marital tensions might well reflect those tensions in his relationships with others, including his children. Furthermore, it is probable that there is a greater incidence of broken homes in the background of problem children than in that of non-problem children. Consequently, it is doubtful if matching the two groups of mothers on this particular variable would be other than an artificial procedure. It seems logical to hypothesize that the same factors making for those parental attitudes assumed to be associated with children's personality or behavior problems also make for marital tensions. If this is true, then one would expect to find, as is true in the present case, a higher percentage of divorced mothers (or unhappily married mothers) in a sample of those who have problem youngsters as contrasted with a sample of those whose children are classifiable as non-problem.

5. *Psychological Help*

In the problem group, two mothers reported that they had undergone psychiatric treatment or psychological counseling, one indicated that her

spouse had had treatment of a psychological sort, and two indicated that both they and their husbands had been exposed to psychotherapy. In the non-problem group, two mothers indicated that both they and their husbands had availed themselves of treatment, and one reported that her husband had had therapy. This seems to be a rather even distribution in the two groups and it seems reasonable to conclude that the responses of the two sets of mothers will be relatively uninfluenced differentially by contact with a psychotherapist.

6. *Conclusions with Respect to the Groups*

The original scale was administered to a group of 100 white, urban mothers, 50 in the problem group and 50 composing a non-problem group. When compared for socio-economic factors, the two groups showed no appreciable differences. The same was true for exposure to psychological counseling or treatment. On the variable of marital stability, the problem showed twice as many divorces as did the non-problem group; this, however, is rationalized on the ground that it is entirely possible that the same factors producing parental attitudes hypothetically correlated with childhood maladjustment are also associated with marital unhappiness. Consequently, one would expect to find a higher incidence of divorces in the problem as opposed to the non-problem group.

On the matter of age, the problem group is very significantly older, but the psychological importance of this is not clear. To see whether this might reflect a significantly greater age in the problem children, which might mean that these youngsters had simply lived longer and consequently had more time in which to develop maladjustments, the mean ages of the problem and non-problem children were found and evaluated comparatively in terms of the *t*-test. It was found that the slight difference in the direction of greater age among the problem children was highly insignificant.

One variable of considerable importance, however, is decidedly uncontrolled in the two groups. In terms of the educational attainments as measured in terms of years of school attended by both the mothers and their husbands, there are significant differences in the direction of greater education in the non-problem group. This uncontrolled variable introduces spurious differences into the responses of the two groups and must be borne in mind when the Survey is evaluated.

The next section will present the results of the item analysis done after the original scale was administered and will describe the procedures followed from that point to the determination of the Survey's reliability and validity.

IV. THE U. S. C. PARENT ATTITUDE SURVEY: THE RELIABILITY AND VALIDITY OF THE TENTATIVE SCALE

This section will describe the procedures followed in dealing with the data gathered from the administration of the original version of the U. S. C. Parent Attitude Survey to the groups described in the preceding section. The particular topics covered will be (*a*) the derivation of the items composing the tentative scale,[6] (*b*) the variables of the Survey and their interrelationships, (*c*) the reliability of the total scale and its variables, and (*d*) the validity of the scale as determined by various techniques.

A. The Derivation of the Tentative Scale

After the original scale was given to the sample indicated in the preceding chapter, an item analysis was undertaken to determine which items in the original 148 possessed sufficient discriminatory value to be retained. The method followed was to make a chi-square test of significance between the responses of the non-problem and problem subjects to each item. Items were retained if they discriminated at the 5 per cent level of confidence or beyond.

This analysis of the original scale items yielded a total of 85 items meeting the test of discrimination at the 5 per cent level or better. These were then weighted according to Guilford's formula (14),

$$IV = \frac{Pp - Pnp}{pq} + 4$$

where

Pp = proportion of the problem group responding in a specified way;
Pnp = proportion of the non-problem group responding in the same way;
p = proportion of the two groups combined according to the formula $p = \frac{Pp + Pnp}{2}$;
$q = 1 - p$.

The resulting weights for each of the 85 retained items are given in the appendix.

[6]The term "tentative scale" as here used, it will be remembered, refers to the version of the Survey derived from the analysis of the results obtained when the original scale was administered to the sample previously described. The term carries the necessary implication that the Survey *is not to be regarded as in final* form as a result of a study of 100 cases.

A word of defense is in order in relation to the procedure of weighting. It was resorted to in this instance in order to take into fullest account the four scoring categories of *Strongly Agree, Mildly Agree, Mildly Disagree,* and *Strongly Disagree* as they differentially contribute to the separation of problem from non-problem parents. To score responses one or zero only would be, in this case, to loose much of the significance of the items. In some cases, for example, differences between the groups show up in terms of the problem subjects' favoring the extremes of the response categories, whereas the non-problem subjects consistently choose the middle ones. How much the gain resulting from weighting is actual and how much a matter of magnifying chance errors in the sample studied will be discussed later when the problem of validity is taken up with special reference to shrinkage.

The tentative version of the U. S. C. Parent Attitude Survey, then, consists of 85 items selected on the basis of their having differentiated between problem and non-problem parents at the 5 per cent level of confidence or beyond, and weighted according to the differential contribution to discrimination of each of the four response categories.

B. The Variables of the Survey and Their Interrelationships

While the chief hypothesis underlying this research is that parent behavior bears sufficient relationship to child adjustment to permit the measurement of the former against the criterion of the latter, the demand for clinical usefulness in such a measure necessitates a step beyond the mere separation of problem parents from non-problem parents. Complying with this necessity, an attempt has been made to determine the variables or sub-scales which compose such a complex entity as parental attitudes toward children. This task has been undertaken to answer the question of what particular kinds of attitudes are of most importance if and when it is shown that parental attitudes are of etiological significance for some child's maladjustment.

Several methodological difficulties arise at this point. Ordinarily, the process of test construction involves the prior establishment of sub-scales which are then validated against some outside criterion. Here, however, one of the main objectives of the study is to collect a pool of items which differentiate on whatever grounds the parents of problem children from those of non-problem children. This objective, of course, is a direct and practical outgrowth of the major hypothesis underlying the research and demands no apology; but it is somewhat at variance with the idea of several independent criteria against which to evaluate sub-scales, since the retention of all items in the Survey has been dependent on the single criterion of how well each predicts membership in a class.

In a situation like this a procedure which may be resorted to is that of factor analysis. An obstacle standing in the way of the use of this method here is the tremendous amount of labor it requires. In the present study a factor analysis based on the intercorrelations among the 85 items in the tentative scale would involve a matrix of 3,570 correlation coefficients. Such computational exercise would hardly be justified on the basis of the relatively small sample on which the investigation is based. Therefore, factor analysis is rejected not only because of the prohibitive amount of time and labor it would require, but more legitimately because of the dubiousness of the results which would accrue from such a procedure.

The line of attack adopted here in an effort to extract significant although tentatively held variables or sub-scales is a variant of what Champney (8) has called "armchair factor analysis," augmented by the use of pooled judgments. The operations performed were as follows: Each of the 85 discriminating items, typed on a card, was considered in the light of the distribution of problem as opposed to non-problem responses to it and in relation to the question, "What seems to be the pattern of the parent-child relationship indicated by this item *as it has been answered?*" On the basis of this procedure, the writer classified the items into three categories plus a small miscellaneous group of 10 items which seemed to defy classification. The sub-scales were named according to the attitudinal themes which the respective sets of items seemed to have in common. The *Dominant* variable, for example, consists of items reflecting a tendency on the part of the parent to put the child in a subordinate rôle, to take him into account quite fully but always as one who should conform completely to parental wishes under penalty of severe punishment. The *Possessive* sub-scale refers to a tendency on the part of the parent to "baby" the child, to emphasize unduly (from a mental hygiene point of view) the affectional bonds between parent and child, to value highly the child's dependence on the parent, and to restrict the child's activities to those which can be carried on in his own family group. The third sub-scale, called the *Ignoring* variable, refers to a tendency on the part of the parent to disregard the child as an individual member of the family, to regard the "good" child as the one who demands the least parental time, and to disclaim responsibility for the child's behavior.

To check on this attempt to establish the relevant sub-scales without recourse to outside criteria, the following procedure was used. Four graduate students, all of whom had M.A. degrees in clinical psychology and were Ph.D. candidates, were given (*a*) slips on which were written the definitions given above of the three sub-scales and (*b*) a packet of 85 cards, on each of which was written one of the significant items of the tentative scale to-

gether with the distribution of problem and non-problem responses to it. Four boxes were provided, labeled *Dominant, Possessive, Ignoring,* and *Miscellaneous.* The directions used read:

> You are to sort the items that you have into the classifications described on your definition sheets. Read each item carefully and pay particular attention to the way in which it was responded to by the parents of problem children and those of non-problem children. If, on the basis of your consideration of the item and the responses to it, you believe it belongs in the Dominant category, put it in the box marked Dominant. If you believe it belongs in the Possessive category, put it in the correspondingly marked box, and so on. If, after proper consideration, you are unable to classify the item, put it in the Miscellaneous box.

The percentages of agreement among the judgments of the four graduate students and the writer are presented in Table 5. Items were retained in any given category when they were placed there by three or more of

TABLE 5
PERCENTAGE AGREEMENT AMONG JUDGES ASSIGNING ITEMS TO SUB-SCALES

Judges	Dominant	Possessive	Ignoring	Miscellaneous
A-B	84	79	88	92
A-C	87	82	91	86
A-D	82	89	86	89
A-E	93	90	91	94
B-C	87	94	88	84
B-D	79	91	88	83
B-E	91	79	78	89
C-D	84	78	91	82
C-E	78	84	89	90
D-E	90	88	93	79
Sum	855	854	883	868
Mean	85.5	85.4	88.3	86.8
SD	4.74	5.44	3.95	4.53

the judges. It is felt that the degree of agreement achieved is sufficient to warrant the tentative use of the sub-scales so derived.

A further check on the usefulness of these sub-scales was made by determining their independence, that is, by measuring the degree to which they intercorrelated and thus overlapped in their contribution to the variance in the total tentative scale from which they were drawn. The obtained intercorrelations are given in Table 6. The degree of independence here reflected is sufficient to permit one to look upon the three sub-scales as variables of the complex totality measured by the tentative scale as a whole; in other words, it seems legitimate to regard the sub-scales tentatively as having sufficient homogeneity within themselves and sufficient independence

among themselves to be useful in assessing the component parts of the complex of attitudes involved in the Survey.

TABLE 6
INTERCORRELATIONS OF SCALES IN U. S. C. PARENT ATTITUDE SURVEY

	Total	Dominant	Possessive	Ignoring
Total	—	.86	.77	.68
Dominant	.86	—	.48	.37
Possessive	.77	.48	—	.40
Ignoring	.68	.37	.40	—

C. THE RELIABILITY OF THE TENTATIVE SCALE AND SUB-SCALES

The reliability of the Survey was determined by the split-half method, raised by the Spearman-Brown formula. The reliability coefficients are given in Table 7. The reliability values are all of sufficient magnitude to permit the interpretation of a high degree of consistency in the Survey.

TABLE 7
RELIABILITY OF TOTAL AND SUB-SCALES
(Raised by Spearman-Brown)

Modality	*r*
Total Scale	.95
Dominant	.91
Possessive	.90
Ignoring	.84

No attempt was made to check reliability by means of the test-retest method because parental attitudes may change as a function of time, and it is difficult to know whether the obtained correlations reflect inconsistency in the instrument or in the subjects measured.

D. THE VALIDITY OF THE TENTATIVE SCALE

The crucial problem to be met in all attempts at measurement is that of validity or the usefulness of the measuring instrument. In this case, the criterion of usefulness is how well the U. S. C. Parent Attitude Survey discriminates between the parents of problem children and those of non-problem children.

It will be recalled that in building the pool of items that currently constitutes the tentative scale, the original scale was administered to a group of 100 mothers, 50 of whom were the parents of problem children while 50 were parents of non-problem youngsters; the sample was characterized by the various points made in Section III. Following this administration, the items were analyzed by the chi-square method to find those which significantly discriminated, and these were retained in the tentative scale.

TABLE 8

RANGES, MEANS AND STANDARD DEVIATIONS OF SCORES OF ORIGINAL AND NEW SAMPLES TO WHICH SURVEY GIVEN

Survey variable	Total group ($N = 100$) Range	M		Problem group ($N = 50$) Range	M		Non-problem group ($N = 50$) Range	M	
				Original sample					
Total Scale	265-466	361.18	40.25	301-466	397.71	29.44	265-384	324.65	21.65
Dominant	*125-237*	*171.25*	*20.55*	*146-237*	*187.82*	*16.73*	125-177	154.68	11.56
Possessive	73-127	101.40	13.09	80-127	111.74	10.27	73- 96	91.06	6.43
Ignoring	30- 67	49.47	6.89	39- 67	56.14	5.85	30- 51	42.80	4.20
				New sample					
	($N = 40$)			($N = 20$)			($N = 20$)		
Total Scale	281-460	374.92	52.40	329-460	399.83	23.49	281-400	350.01	20.93
Dominant	135-240	189.61	18.59	154-240	201.19	14.10	135-217	178.03	12.64
Possessive	69-138	103.19	12.57	74-138	113.26	9.41	69-107	94.12	7.90
Ignoring	30- 74	52.22	7.19	36- 74	56.71	6.08	30- 69	47.73	6.49

The first attempt at validation involves the presentation of validity coefficients derived from this first administration. While it is true that to evaluate a newly developed test against the samples which it was specially designed to discriminate is merely to magnify chance error and hence spuriously to inflate correlations, such a step is worth taking if only in order to determine the amount of shrinkage in the validity coefficient from the original to a new sample. Therefore, the data on the original administration are given in Table 8, where the ranges, means, and standard deviations are shown for the total sample and its problem and non-problem halves. The validity coefficients based on these data are presented in Table 9.

The point-biserial coefficient of correlation (24) was used because, while the scores on the Survey distributed themselves in a way closely approximating

TABLE 9
VALIDITY COEFFICIENTS FROM ORIGINAL ADMINISTRATION OF SURVEY
(N = 50 Problem Parents + 50 Non-Problem)

Survey Variable	Point-Biserial r
Total Scale	.904
Dominant	.801
Possessive	.790
Ignoring	.968

normality, the legitimacy of assuming a normal distribution in child adjustment seemed highly questionable. While there is little doubt that such a variable can be so defined that it is continuously distributed, it is a moot topic as to whether it would be normally distributed (and thus fulfill the conditions necessary for using a biserial r). Certainly, the criteria used in this study for determining child adjustment are rather clearly dichotomous. A child has either received clinical attention, or he has not; he either has a record of two or more contacts with the juvenile authorities, or he has not; his mother has either sought clinical aid for him, or she has not. Consequently, the point-biserial was resorted to on the assumption that one variable was continuously and normally distributed whereas the other was a true point distribution.

Following its original administration, the Survey was given to 40 mothers, again divided equally between the problem and non-problem categories. No attempt was made to obtain background information on these women aside from that necessary to classify them correctly from the point of view of the investigation. The data taken from this second administration are also presented in terms of ranges, means, and standard deviations in Table 8.

TABLE 10
VALIDITY COEFFICIENTS FROM SECOND ADMINISTRATION OF SURVEY
(N = 20 Problem Parents + 20 Non-Problem)

Survey Variable	Point-Biserial r
Total Scale	.769
Dominant	.623
Possessive	.721
Ignoring	.624

The correlation coefficients drawn from these data are given in Table 10.

It is obvious from a comparison of Tables 9 and 10 that there has been a sharp shrinkage in the magnitude of the correlation coefficients which serve as indices of the Survey's validity. The shrinkages are summarized in Table 11.

TABLE 11
SHRINKAGE IN VALIDITY COEFFICIENTS FROM ORIGINAL TO SECOND ADMINISTRATIONS

Survey Variable	Original	Second	Shrinkage
Total Scale	.904	.769	.135
Dominant	.801	.623	.178
Possessive	.790	.721	.069
Ignoring	.968	.624	.344

The amount of shrinkage occurring is not excessive, however, and the validity measures obtained from the second administration are still quite high. This permits the tentative interpretation that the Survey has some genuine relevance in the assessment of parent attitudes in relation to child adjustment.

This, of course, bears out the hypotheses on the basis of which the study was begun. The construction of a test which differentiates the criterion groups here used implies clearly (*a*) that parents behave with identifiable consistency toward their children and (*b*) that parental attitudes are significantly related to child adjustment. As a matter of fact, if the coefficient of determination (r^2) is computed from the validity coefficients on the new sample of mothers, it becomes apparent that roughly half the variance in the criterion of child adjustment is predicted by the parents' scores on the Survey.

A second attempt to check on the validity of the Survey was made in terms of the comparison of scores obtained by parents with scores obtained by clinical psychologists, where the latter were expected to provide a kind of ideal which would be more closely approximated by the parents of non-problem children than by those with problem youngsters. With this hypothe-

sis in mind, 10 clinical psychologists, all of them Ph.D.'s, were asked to take the Survey under the following directions: "In the light of your knowledge of the mental hygiene of parent-child relationships, please fill out the attached questionnaire in the manner which you feel would characterize an ideal parent." Eight of the psychologists contacted responded. While this is certainly an inconsequential number of cases for most purposes, it seems legitimate in this connection because the only aim in using these clinicians was to set up an ideal. The scores are summarized in Table 12.

TABLE 12
"IDEAL SCORES" OF CLINICAL PSYCHOLOGISTS
($N = 8$)

Survey Variable	Range	Mean	*SD*
Total Scale	275-297	286.38	6.67
Dominant	142-149	146.12	2.21
Possessive	72-88	79.38	4.62
Ignoring	35-44	39.88	2.72

From the narrow ranges and small standard deviations, it is readily apparent that the clinicians agreed to a marked extent on the responses of an "ideal" parent.

These "ideal" scores were then compared with the scores made by the problem and non-problem groups to which the Survey was first administered. The *t*-test was used. The results are given in Table 13.

TABLE 13
SIGNIFICANCE OF DIFFERENCES BETWEEN CLINICIANS' SCORES AND ORIGINAL PARENTS' SCORES

Survey Variable	*Clinicians* vs. Problem group	*Clinicians* vs. Non-Problem group
Total Scale	$t = 22.69$	$t = 9.56$
Dominant	16.30	4.62
Possessive	13.75	5.92
Ignoring	9.26	2.06

Inspection of the table of *t*-values above reveals that both parent groups differ in their responses from the clinicians beyond the 1 per cent level of confidence in all cases except one. The one instance shows a difference at the 5 per cent level only between clinicians and non-problem parents on the *Ignoring* sub-scale. The important thing from the point of view of this study is that without exception the differences between the "ideal" clinicians' scores and those of the problem group are much greater than those between the clinicians and the non-problem group. This may be taken as a further index,

albeit a rough one, of validity in the sense that it bears out the prediction that mothers of problem children will depart more drastically in their attitudes toward their youngsters from a mental hygiene norm of parental attitudes than will the mothers of non-problem children.

A third attempt at validation proved abortive. Ten four-point rating scales, based on the variables of the Fels Parent-Behavior Rating Scales were constructed and distributed to the various clinicians and agencies from which the problem group of parents was drawn. Clinicians were asked to rate the parents independently of their responses to the U. S. C. Parent Attitude Survey in order to provide an objective check on the parent-child relationship as understood by a clinical worker intimately acquainted with the situation. The plan had to be abandoned because the raters felt that they were too busy to do the job. The number of returned rating sheets was much too low to permit their being used in the study.

The tentative scale as now established, showing the apportionment of items to the three sub-scales, the items dropped from the original scale, and the weights for the retained items, is presented in the appendix.

V. SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

The hypothesis on which this investigation has been based is that parents take sufficiently consistent attitudes toward their children to permit measurement and that those attitudes are significantly related to child adjustment. By the construction of the U. S. C. Parent Attitude Survey, an inventory-type "test" of parent attitudes toward their children, which has sufficient reliability and validity to be clinically useful, the hypothesis may be looked upon as at least tentatively substantiated.

The procedure was to comb the literature on parent-child relationships for hints as to items that would successfully differentiate the parents of problem children from those of non-problem children. Problem children were defined as those who had come afoul of the juvenile authorities at least twice, who were undergoing clinical treatment for some personality or behavior disorder, or who had been complained about by the mother as constituting a problem for which she would like to have aid. The non-problem children were, of course, those who did not meet any of these three criteria of maladjustment.

A pool of 148 items was formed and administered to a group of 50 mothers of problem children and a group of 50 mothers of non-problem children. On the basis of this preliminary administration, the items were analyzed for significance by the chi-square method, those items being kept which differentiated the groups at the 5 per cent level of confidence or better. Eighty-five items were retained.

Sub-scales were then extracted by having five sophisticated judges classify the items according to the categories *Dominant, Possessive, Ignoring,* and *Miscellaneous*. On the basis of the agreement among the judges, 10 items which defied classification were kept in the survey. The others were grouped in sub-scales as named. The *Dominant* variable consists of items reflecting a tendency on the part of the parent to put the child in a subordinate rôle, to take him into account quite fully but always as one who should conform completely to parental wishes under penalty of severe punishment. The *Possessive* sub-scale refers to a tendency to "baby" the child, to emphasize unduly the affectional bonds between parent and child, to value highly the child's dependence on the parent, and to restrict the child's activities to those which can be carried on in his own family group. The *Ignoring* sub-scale refers to a tendency on the part of the parent to disregard the child as an individual member of the family, to regard the "good" child as the one who demands the least parental time, and to disclaim responsibility for the child's behavior.

Validity coefficients were computed for the original group and then for a new group of 20 mothers of problem children and 20 mothers of non-problem children. The procedure was undertaken to estimate the shrinkage that is inevitable when a predictive device built on one sample is applied to a new and different sample. Shrinkage was surprisingly small, and the validities on the new group were as follows: Total Scale, .769; Dominant, .623; Possessive, .721; and Ignoring, .624. This means that roughly half the variance in the criterion of child adjustment may be predicted from the attitude scores. Reliability in the total and three sub-scales was high.

On this basis of this procedure we may conclude that: (*a*) Parent behavior, as represented by parental attitudes, is measurably consistent; that (*b*) parent attitudes are meaningfully associated with child adjustment; that (*c*) apparently relevant and internally consistent variables can be extracted from a pool of items by means of the combined judgments of sophisticated judges; and that (*d*) the U. S. C. Parent Attitude Survey, an easily administered pencil-and-paper type of inventory, is of sizeable potential value in the investigation of parental attitudes as they affect children's adjustment.

A number of suggestions for further work immediately suggest themselves, a few of which may be mentioned here. In the first place, the U. S. C. Parent Attitude Survey gives promise of being a useful instrument and needs to have norms established on a large random sample at the earliest possible date. Second, the problem of validity needs to be further explored in terms of such outside criteria as clinical case reports, home visits, and/or pooled ratings by those who know the subjects well. Third, the derivation of sub-scales should be checked with reference to the correspondence between sub-scale scores and objective parent-child relationship factors. Fourth, the sub-scales could probably be purified of the degree of overlap that now exists if a matrix of item intercorrelations could be established and some form of cluster analysis applied.

Beyond the Survey itself, of course, lie a number of problems in the solving of which it could be a highly useful instrument once it has been sufficiently refined. Among other questions are such as these: Are specific patterns of parent behavior associated with specific types of childhood adjustment problems? Do the attitudes of mothers and of fathers differ markedly, and what is the relationship between such possible differences and childhood adjustment patterns? What parental attitudes associated with childhood maladjustment are most amenable to change and through what therapeutic processes? What are the relationships between parental attitudes and certain factors like intelligence, socio-economic background, degree of education, and the parent's own experience as a child?

APPENDIX: UNIVERSITY OF SOUTHERN CALIFORNIA PARENT ATTITUDE SURVEY

Read each of the statements below. Rate each statement as to whether you *strongly agree, mildly agree, mildly disagree* or *strongly disagree*. There are no right or wrong answers, so answer according to your own convictions. Work as rapidly as you can. Draw a circle around the letter that best expresses your feeling.

Variable		Item	Weights: Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
Ignoring	1.	A child should be seen and not heard.	6	5	4	3
Possessive	2.	Parents should sacrifice everything for their children.	6	5	3	4
	3.	A child should be required to consult his parents before making any kind of decision.	Dropped			
Possessive	4.	Children should be allowed to do as they please.	5	5	2	5
	5.	Children should feel free to bring their playmates into the house or yard to play.	Dropped			
	6.	No child should be given so much freedom that he will begin to feel independent.	Dropped			
	7.	A child should not plan to enter any occupation his parents don't approve of.	6	6	4	5
	8.	Parents should not compel a child to do things which cause him great embarrassment.	Dropped			
	9.	Spanking is a good form of discipline.	Dropped			
Dominating	10.	Children need some of the natural meanness taken out of them.	6	5	3	4
Dominating	11.	A child should have strict discipline in order to develop a fine, strong character.	6	4	3	3
Ignoring	12.	The mother rather than the father should be responsible for discipline.	6	4	3	4
Possessive	13.	Children should be "babied" until they are several years old.	6	5	3	4
Dominating	14.	Children have the right to play with whomever they like.	4	3	5	5
	15.	Parents should realize that they teach the child bad habits as well as good ones.	Dropped			
	16.	Children who are bold and daring are more fun than those who are careful and cautious.	Dropped			
Possessive	17.	Independent and mature children are less lovable than those children who openly and obviously want and need their parents.	6	5	4	3
	18.	Children should be expected to buy certain necessities out of their allowances.	Dropped			
	19.	Parents should obtain obedience by appealing to the child's better self.	Dropped			
	20.	Children, after all, behave just as they are taught to behave.	Dropped			

Variable		Item	Weights: Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
Possessive	21.	Children should be forbidden to play with youngsters whom their parents do not approve of.	5	5	3	2
Ignoring	22.	A good way to discipline a child is to tell him his parents won't love him any more if he is bad.	6	3	4	4
Dominating	23.	Severe discipline is essential in the training of children.	6	6	5	3
	24.	Fathers rather than mothers should administer the discipline in the home.	Dropped			
Ignoring	25.	Parents cannot help it if their children are naughty.	6	5	4	3
Unclassified	26.	Jealousy among brothers and sisters is a very unhealthy thing.	4	5	2	6
Dominating	27.	Children should be allowed to go to any Sunday School their friends go to.	5	2	4	5
	28.	Children ordinarily quit trying to get their own way and do what their parents say if the parent is always firm.	Dropped			
	29.	Parents should not be blamed for the misbehavior of their children.	Dropped			
Dominating	30.	No child should ever set his will against that of his parents.	6	6	2	4
	31.	The best parents discard authority for friendship in dealing with their children.	Dropped			
	32.	Parents should always treat their children with courtesy and respect.	Dropped			
Dominating	33.	The Biblical command that children must obey their parents should be completely adhered to.	6	4	4	3
	34.	A child should be given money only when he is helpful and obedient.	Dropped			
	35.	Children usually have temper tantrums if they can't have their own way.	Dropped			
Dominating	36.	It is wicked for children to disobey their parents.	6	4	4	3
	37.	Children should not be permitted to learn about sex until they are old enough to understand all about it.	Dropped			
	38.	Children generally tend to pout and sulk when they can't have their own way.	Dropped			
	39.	Chidren should follow the family profession or occupation.	Dropped			
Possessive	40.	A child should feel a deep sense of obligation always to act in accord with the wishes of his parents.	6	5	3	3
Possessive	41.	Children should not be punished for disobedience.	5	6	3	4
	42.	Children should be allowed the time to talk over with parents what is best.	Dropped			

Variable		Item	Weights: Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
Possessive	43.	Children who are gentlemanly or ladylike are preferable to those who are tomboys or "regular guys."	5	5	3	3
	44.	Children should be urged to develop themselves to the utmost—to get ahead as rapidly and as far as possible.	Dropped			
Dominating	45.	Strict discipline weakens a child's personality.	4	3	4	5
	46.	Making a child suffer the natural consequences of his bad behavior is a good form of discipline.	Dropped			
Possessive	47.	Children should always be loyal to their parents above anyone else.	6	3	4	3
Dominating	48.	Children should be steered away from the temptations of religious beliefs other than those accepted by the family.	6	6	3	3
	49.	Children should have a say in the making of family plans.	Dropped			
Possessive	50.	The weaning of a child from the emotional ties to its parents begins at birth.	5	3	4	5
Possessive	51.	Parents are not entitled to the love of their children unless they earn it.	4	3	5	6
	52.	It is the duty of parents to encourage their children to accomplish more than they themselves did.	Dropped			
Dominating	53.	Parents should never try to break a child's will.	4	2	5	5
Dominating	54.	Children should not be required to take orders from parents.	2	5	4	5
Dominating	55.	Children should be allowed to choose their own religious beliefs.	4	3	4	6
Ignoring	56.	Chidren should not interrupt adult conversation.	5	4	2	6
Ignoring	57.	The most important consideration in planning the activities of the home should be the needs and interests of children.	4	2	5	6
	58.	Quiet children are much nicer than little chatter-boxes.	6	4	3	4
Ignoring	59.	It is sometimes necessary for the parent to break the child's will.	6	5	4	3
Dominating	60.	Children who get along best with other children are nicer than children who are more popular with adults.	Dropped			
	61.	A good way to discipline a child is to shame him before his playmates.	Dropped			
Unclassified	62.	Children usually know ahead of time whether or not parents will punish them for their actions.	5	3	3	4

Variable		Item	Weights			
			Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
Dominating	63.	Children resent discipline.	5	4	3	5
	64.	The best way to get a child to obey is to warn him that God will punish him (or the bogeyman will get him, etc.).	Dropped			
Possessive	65.	Children should not be permitted to play with youngsters from the "wrong side of the tracks."	6	5	3	4
	66.	Parents should always be satisfied with the way their own children act.	Dropped			
	67.	Problem children are usually jealous of their brothers and sisters.	Dropped			
Dominating	68.	When the parent speaks, the child should obey.	5	5	3	2
	69.	Children should be neat and stay within calling distance of their parents.	Dropped			
Dominating	70.	Mild discipline is best.	4	3	5	6
	71.	Children should be shown lots of affection.	Dropped			
	72.	Parents have no right to try to influence a child's choice of occupation.	Dropped			
	73.	Children should be offered rewards to get them to obey.	Dropped			
Possessive	74.	The best child is one who shows lots of affection for his mother.	6	5	3	4
	75.	Little arguments and fights that break out among children are best handled by the children themselves.	Dropped			
Dominating	76.	A child should be taught that his parents always know what is best.	5	5	3	3
	77.	Parents should watch their children to protect them from harm.	Dropped			
	78.	Any sign of jealousy in a child should be punished.	Dropped			
	79.	Children should be disciplined chiefly by taking away privileges.	Dropped			
Possessive	80.	It is better for children to play at home than to visit other children.	6	4	4	3
	81.	Children should be allowed to play only with their social equals.	Dropped			
Dominating	82.	Most children should have more discipline than they get.	6	4	3	2
	83.	Parents ought to close their eyes to their children's faults.	Dropped			
	84.	Parents should make all important decisions for their children.	Dropped			
	85.	A child will resent too much love and petting from his parents.	Dropped			
Dominating	86.	A child should do what he is told to do, without stopping to argue about it.	6	4	3	4
Dominating	87.	Children should fear their parents to some degree.	6	5	4	3

Variable		Item	Weights: Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
	88.	Children most often get their own way by ignoring parents and doing what they please.	Dropped			
Possessive	89.	A child should always love his parents above everyone else.	6	+	3	+
Unclassified	90.	Children who indulge in sex play become adult sex criminals.	5	6	+	3
	91.	Isolation (sending a child off to be alone or making him stay in his room) is a good form of discipline.	Dropped			
Dominating	92.	Children should be allowed to make only minor decisions for themselves.	5	5	3	3
Dominating	93.	A child should always accept the decision of his parents.	5	5	3	3
	94.	Masturbation in children should generally be ignored by the parents.	Dropped			
	95.	Before disciplining a child, a parent should carefully explain why he is being punished.	Dropped			
Possessive	96.	Children who readily accept authority are much nicer than those who try to be dominant themselves.	6	+	3	3
Dominating	97.	Parents should always have complete control over the actions of their children.	5	+	+	3
	98.	Parents should train their children so they will always be eager to help.	Dropped			
Ignoring	99.	When they can't have their own way, children usually try to bargain or reason with parents.	5	3	+	6
Unclassified	100.	The shy child is worse off than the one who masturbates.	+	3	5	5
	101.	Brothers and sisters are usually jealous of the extra attention given problem children.	*Dropped*			
	102.	No child should be forced to adopt any religious creed to which he objects.	Dropped			
	103.	Children would rather be spanked than told that their mothers won't like them any more.	Dropped			
Dominating	104.	Children should accept the religion of their parents without question.	5	6	+	3
Dominating	105.	The child should not question the commands of his parents.	6	+	3	3
Ignoring	106.	Children who fight with their brothers and sisters are generally a source of great irritation and annoyance to their parents.	6	3	+	6
	107.	A child should have an allowance to be spent without consulting his parents.	Dropped			

Variable		Item	Weights			
			Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
Dominating	108.	Children should not be punished for doing anything they have seen their parents do.	+	+	3	6
Ignoring	109.	Jealousy is just a sign of selfishness.	6	3	+	+
Unclassified	110.	Children should be taught the value of money early.	5	3	3	6
Dominating	111.	A child should be punished for contradicting his parents.	6	5	3	3
Dominating	112.	Children should have lots of parental supervision.	5	3	3	+
Possessive	113.	A parent should see to it that his child plays only with the right kind of children.	6	+	3	3
Possessive	114.	Babies are more fun for parents than older children are.	6	5	+	3
Possessive	115.	Parents should supervise a child's selection of playmates very carefully.	6	+	2	+
Unclassified	116.	No one should expect a child to respect parents who nag and scold.	5	3	5	2
Dominating	117.	A child should always believe what his parents tell him.	6	+	3	+
Dominating	118.	Children should usually be allowed to have their own way.	6	5	3	6
	119.	Parents should spend as much time as possible playing with their children.	Dropped			
	120.	Children who are well liked by their brothers and sisters are often more difficult to manage than those who are not.	Dropped			
Unclassified	121.	A good way to discipline a child is to cut down his allowance.	5	+	3	+
Dominating	122.	Children should not be coaxed or petted into obedience.	+	3	6	5
	123.	Children should not be allowed to go to public playgrounds to play.	Dropped			
Dominating	124.	A child should be shamed into obedience if he won't listen to reason.	6	5	+	+
Possessive	125.	In the long run it is better, after all, for a child to be kept fairly close to his mother's apron strings.	6	6	3	3
Dominating	126.	A good whipping now and then never hurt any child.	6	+	3	2
Unclassified	127.	Masturbation is the worst bad habit that a child can form.	6	5	+	3
Possessive	128.	A child should never keep a secret from his parents.	7	+	3	+
Ignoring	129.	Parents are generally too busy to answer all a child's questions.	6	+	3	3
Dominating	130.	The children who make the best adults are those who obey all the time.	6	5	3	+
Unclassified	131.	It is important for children to have some kind of religious upbringing.	6	3	2	2

Variable		Item	Weights: Strongly agree	Mildly agree	Mildly disagree	Strongly disagree
Ignoring	132.	Children should be allowed to manage their affairs with little supervision from adults.	5	3	4	5
	133.	Children will ordinarily tell their parents their thoughts and feelings, their hopes and fears.	Dropped			
Dominating	134.	Parents should never enter a child's room without permission.	3	3	3	7
Dominating	135.	It is best to give children the impression that parents have no faults.	6	5	4	3
	136.	Children should never be permitted to see their parents undressed.	Dropped			
	137.	If a child does not want to do what he *is asked, a parent should* give in to him.	Dropped			
	138.	A child should have freedom in order to develop a personality of his own.	Dropped			
Ignoring	139.	Children should not annoy their parents with their *unimportant* problems.	6	5	3	4
Dominating	140.	Children should give their parents unquestioning obedience.	6	4	4	2
	141.	All of a child's questions should be answered frankly and patiently.	Dropped			
	142.	Parents should respect their children's desire for privacy, and *if possible* provide them with a place where they can be alone when they want to be.	Dropped			
Unclassified	143.	Sex is one of the greatest problems to be contended with in children.	6	4	3	4
Ignoring	144.	Children should have as much freedom as their parents allow themselves.	6	4	3	6
	145.	Children should be trained from the cradle in habits of independent thought and action.	Dropped			
	146.	Most children have a need for privacy.	Dropped			
Dominating	147.	Children should do nothing without the consent of their parents.	6	5	3	3
	148.	Jealousy among brothers and sisters is a very common thing.	Dropped			

25. RADKE, M. J. The Relation of Parental Authority to Children's Behavior and Attitudes. Minneapolis: Univ. Minnesota Press, 1946.

26. ROSS, B. M. Some traits associated with sibling jealousy in problem children. *Smith Coll. Stud. Soc. Wk.*, 1931, **1**, 364-376.

27. SEWALL, M. Some causes of jealousy in young children. *Smith Coll. Stud. Soc. Wk.*, 1930, **1**, 16-22.

28. STOGDILL, R. Attitudes of parents, students, and mental hygienists toward children's behavior. *J. Soc. Psychol.*, 1933, **4**, 486-489.

29. STRECKER, E. A. Their Mothers' Sons. New York: Lippincott, 1946.

30. SYMONDS, P. The Psychology of Parent-Child Relations. New York: Appleton-Century, 1939.

31. TEAGARDEN, F. Child Psychology for Professional Workers. New York: Prentice-Hall, 1940.

32. UPDEGRAFF, R. Recent approaches to the study of the preschool child: III. Influence of parent attitudes on child behavior. *J. Consult. Psychol.*, 1939, **3**, 34-36.

33. WATSON, G. A comparison of the effects of lax vs. strict home training. *J. Soc. Psychol.*, 1934, **5**, 102-105.

34. ———. A critical note on two attitude studies. *Ment. Hyg.*, 1933, **17**, 59-64.

35. WITMER, H. L. Parental behavior as an index to the probable outcome of treatment in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1933, **3**, 431-434.

36. WYLIE, P. Generation of Vipers. New York: Farrar & Rhinehart, 1942.

) per volume
numbers $4.00

QUARTERLY
Two volumes per year

May, 1949
Volume 39, Second Half

(Founded in 1921 by Carl Murchison)

TIC PSYCHOLOGY

PHS

havior
logy

K. S. Lashley
Harvard University

A. R. Luria
Medico-biological
Institute, Moscow

Henri Piéron
Université de Paris

Robert R. Sears
State University
of Iowa

George D. Stoddard
University of Illinois

Calvin P. Stone
Stanford University

Lewis M. Terman
Stanford University

Godfrey Thomson
University of
Edinburgh

E. L. Thorndike
Teachers College,
Columbia University

C. J. Warden
Columbia University

John B. Watson
New York City

Copyright, 1949, by The Journal Press
Provincetown, Massachusetts, U. S. A.

Entered as second-class matter April 15, 1937, at the post-office at Provincetown, Mass., under Act of March 3, 1879

The Journal of Psychology

EDITED BY

Carl Murchison

The *Journal of Psychology* was founded at the request of many psychologists throughout the world for the purpose of establishing a medium in which publication shall be immediate and in which other publication experiments may be carried out.

(1). The subscription price will be $7 per volume.

(2). The format of each volume will be such that each subscriber may bind down the Assembled Volume into its component articles.

(3). Except through the fault of the author, the lag in publication should not be more than three or four weeks.

(4). The author will receive 200 reprints of his paper, and subscribers will receive the accumulated publications bound into half volume units. The second half of each volume will contain the title page and table of contents for that volume. Each half volume unit will be mailed to the subscribers at its stated date.

(5). Each author will be charged the actual manufacturing costs of any cuts or halftone work.

(6). The publication charges will be as follows:

4	page signature		$32.00
8	"	"	32.00
12	"	"	48.00
16	"	"	64.00
20	"	"	80.00
24	"	"	96.00
28	"	"	112.00
32	"	"	128.00
36	"	"	144.00
40	"	"	160.00

These charges are tentative and will be adjusted as furnished by the subscription list (especially by the effect of the war in China and in Europe).

(7). The accumulated units will be mailed to subscribers *quarterly*: January 5th, July and October. The units may vary at times to accommodate publication needs.

Two volumes annually. Per annum $14.00; per volume $7.00; per half volume unit $4.00. Canadian rate at $7.00 per volume plus transportation.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Single numbers $4.00
$7.00 per volume

QUARTERLY
Two volumes per year

May, 1949
Volume 39, Second Half

(Founded in 1925 by Carl Murchison)

GENETIC PSYCHOLOGY MONOGRAPHS

Child Behavior, Animal Behavior, and Comparative Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

MAY, 1949

(Manuscripts are printed in the order of final acceptance)

QUALITATIVE DIFFERENCES IN THE VOCABULARY RESPONSES OF NORMALS AND ABNORMALS . . 151
By Herman Feifel

THE RELATIVE EFFECTIVENESS OF MOTION AND STILL PICTURES AS STIMULI FOR ELICITING FANTASY STORIES ABOUT ADOLESCENT-PARENT RELATIONSHIPS 207
By Paul E. Eiserer

THE ORGANIZATION OF HEREDITARY MAZE-BRIGHTNESS AND MAZE-DULLNESS . . . 281
By Lloyd V. Searle

Copyright, 1949, by The Journal Press
Provincetown, Massachusetts, U. S. A.
Entered as second-class matter April 15, 1937, at the post-office at Provincetown, Mass., under Act of March 3, 1879

Published as a separate and in *Genetic Psychology Monographs*, 1949, 39, 151-204.

QUALITATIVE DIFFERENCES IN THE VOCABULARY RESPONSES OF NORMALS AND ABNORMALS*[1]

Teachers College, Columbia University

HERMAN FEIFEL

	Acknowledgments	153
I.	Introduction	155
II.	Procedure of investigation	167
III.	Treatment of the data	173
IV.	Summary and conclusions	199
	References	203

*Received in the Editorial Office on November 23, 1948.

[1]Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Faculty of Philosophy, Columbia University, 1948.

ACKNOWLEDGMENTS

The writer expresses deep indebtedness to Professor Irving Lorge for his *constant guidance, and generous coöperation in numerous ways; his* appreciation to Professors Percival M. Symonds and Robert L. Thorndike for their constructive criticism and valuable advice; to Professor Nicholas Hobbs for his helpful suggestions and assistance in securing normal subjects; to Mr. Gordon F. Derner for his help in obtaining normal subjects; to Drs. Isabelle V. Kendig and Margaret Ives for their kind aid in facilitating the gathering of data on abnormal patients of St. Elizabeths Hospital, Washington, D. C.; to Mr. Earl Waller for his coöperation in establishing the reliability of the qualitative categories used in the study; to Mr. Murray Aborn for his help in scoring many of the papers of the normal subjects; to Mrs. Thelma Post for giving freely of her time to assist in various clerical aspects of the data; and deep gratitude to his parents for their devotion.

HERMAN FEIFEL

1547 Pitkin Avenue
Brooklyn 12, New York

I. INTRODUCTION

The important rôle that vocabulary plays in human mental life is well established. The intimate connection between language and the thought processes has been stressed by Piaget (30) who in his work with children pointed out that words are bound up with cognition and are related to concept formation.

A. Review of Literature

1. *Intelligence*

Research studies have long emphasized the close relationship between vocabulary and intelligence. Terman (43), referring to his 1916 Stanford-Binet Scale, stated that "the vocabulary test has a far higher value than any other single test of the scale" and reported a correlation of .91 between mental age and the vocabulary test for 631 children and a correlation of .81 for 482 "miscellaneous adults."

Weisenburg, Roe, and McBride (49) found the same coefficient for an adult group which they considered representative of the average population.

For the new Revised Stanford-Binet Scale, Terman and Merrill observed that the vocabulary test agreed "to a high degree with the mental age rating on the scale as a whole; correlations for single age groups ranging from .65 to .91 with an average of .81" (44, p. 302).

Similar findings were reported by Roe and Shakow (35) who surveyed the relationship in normal and abnormal adults. Using the 1916 Stanford-Binet Scale they found correlations between the vocabulary and the test as a whole, ranging from a .76 in psychoneurosis to .92 in paranoid dementia praecox patients, with the normal group showing a correlation of .81.

2. *Scoring Method*

There are essentially two types of vocabulary tests which are used in intelligence testing. One kind, relying on recall factors, presents a key word to the subject to which he must respond with an acceptable definition. The second variety, emphasizing recognition aspects, introduces a word and then four or five alternative words (multiple-choice technique) from which the subject selects the most appropriate matching one. The scoring system of the recall type assumes that all right answers are of equal value and makes no differentiation between a "good" right answer and a "poor" right answer, except for occasional half-credits, nor does it take into account the different shades, quality, and range of possible meanings.

3. *Age*

Utilizing this "all or none" approach, so to speak, in scoring the recall form of vocabulary test, one of the significant generalizations from the literature on intellectual change during the past three decades has been the apparent stability of vocabulary as compared with other mental capacities after maturity is reached. The observation that vocabulary ability is well maintained throughout life was noted by Terman (43) who found that his 482 "miscellaneous adults" had an average vocabulary score higher than those of children and young adolescents of the same mental age.

Green (21), administering 50 words (45 of which eventually became the 1937 Terman vocabulary) to a group of 110 adults ranging in age from 19 to 84, discovered that there was no great change in vocabulary score after the early twenties were reached.

Gilbert (16), using the Babcock Test of Mental Efficiency, which incorporates the 1916 Terman vocabulary, indicated that vocabulary did not increase materially beyond the twenties.

Shakow and Goldman (37) also employed the 1916 Terman vocabulary with 302 adults. In order to overcome increasing selectivity of subjects of higher intellectual level with increasing age and a resultant increase in vocabulary score spuriously attributed to age, they equalized their age groups for intellectual level by obtaining representative education groups for each decade. Their results showed that vocabulary score remained at practically the same level from age 18 through the sixth decade with a slow decline thereafter.

Roe and Shakow (35), in their work on intelligence in mental disorder, reported that in 69 normal adults the 1916 Terman vocabulary was comparatively little affected by age until late maturity.

Rabin (32, 33), using the 1916 Terman vocabulary with 404 mental patients extending over seven decades of life, and the 1937 Terman vocabulary with 268 psychotic and non-psychotic individuals ranging from age 15 through 70, found that in both instances there was a slight tendency for the vocabulary score to be higher at the latter decades of life.

Other studies throw corroborative light on the age-vocabulary relationship though they do not utilize the Terman vocabulary. Willoughby (50), in a study of parents and children, found little decline with increase in age in vocabulary score when compared with other tests.

Jones and Conrad (24) gave the Army Alpha to 1,191 subjects ranging from age 10 through 60 and established that the Synonym-Antonym (vocabu-

lary) subtest did not show the post-adolescent decline found in all the tests (except the General Information subtest).

Sorenson (39), analyzing results from a group of 641 university extension students ranging up to age 69, reported increasing vocabulary scores for successively older age groups.

Christian and Paterson (9) gave Subtest *A* (vocabulary) of the University of Minnesota College Ability Tests to 329 University of Minnesota freshmen and their relatives. They discovered a steady increase in vocabulary up to the seventies, but pointed out that they were dealing with a superior population and that similar results might not hold for average or below average groups.

Wechsler (48) found that the Vocabulary Test of the Wechsler-Bellevue Scale, even though it did not remain constant right up to senescence, held up very well with age.

Thorndike and Gallup (45), administering an untimed, multiple-choice vocabulary test made up of items taken from the *IER* Intelligence Scale *CAVD* to a representative sample of the American adult voting public, reported that age differences for the test were almost nil, and that it was not until the "over sixty" age group was reached that any substantial drop in vocabulary score was discovered.

Sward (40), matching a superior population of 45 university professors aged 60 to 80 against a control group of 45 younger professors and instructors aged 25 to 35, concluded that in word knowledge or general vocabulary, the senescent are uniformly superior to the younger middle-aged.

Fox (14) matched a group of 30 subjects, aged 70 to 79, against an equivalent group of 30 subjects, aged 40 to 49, using both definition and multiple-choice types of vocabulary tests, and noted no difference in the vocabulary scores of the older group. She also indicated that people, regardless of age, found it more difficult to state the meaning of a word than to choose a synonym from among four alternatives.

The experimental evidence just cited ostensibly substantiates the contention that vocabulary, compared with other test functions, declines little, if at all, with increasing age. However, it is quite possible that these results may be explained by the crude scoring system utilized which tends to obscure characteristic differences in the quality of response in the definition type of vocabulary, and by the fact that in the multiple-choice type of vocabulary the correct answer is suggested even though the meaning of the word may have been forgotten. Since the qualitative aspects of intellectual performance have been acknowledged as being fully as important as its quan-

3. *Age*

Utilizing this "all or none" approach, so to speak, in scoring the recall form of vocabulary test, one of the significant generalizations from the literature on intellectual change during the past three decades has been the apparent stability of vocabulary as compared with other mental capacities after maturity is reached. The observation that vocabulary ability is well maintained throughout life was noted by Terman (43) who found that his 482 "miscellaneous adults" had an average vocabulary score higher than those of children and young adolescents of the same mental age.

Green (21), administering 50 words (45 of which eventually became the 1937 Terman vocabulary) to a group of 110 adults ranging in age from 19 to 84, discovered that there was no great change in vocabulary score after the early twenties were reached.

Gilbert (16), using the Babcock Test of Mental Efficiency, which incorporates the 1916 Terman vocabulary, indicated that vocabulary did not increase materially beyond the twenties.

Shakow and Goldman (37) also employed the 1916 Terman vocabulary with 302 adults. In order to overcome increasing selectivity of subjects of higher intellectual level with increasing age and a resultant increase in vocabulary score spuriously attributed to age, they equalized their age groups for intellectual level by obtaining representative education groups for each decade. Their results showed that vocabulary score remained at practically the same level from age 18 through the sixth decade with a slow decline thereafter.

Roe and Shakow (35), in their work on intelligence in mental disorder, reported that in 69 normal adults the 1916 Terman vocabulary was comparatively little affected by age until late maturity.

Rabin (32, 33), using the 1916 Terman vocabulary with 404 mental patients extending over seven decades of life, and the 1937 Terman vocabulary with 268 psychotic and non-psychotic individuals ranging from age 15 through 70, found that in both instances there was a slight tendency for the vocabulary score to be higher at the latter decades of life.

Other studies throw corroborative light on the age-vocabulary relationship though they do not utilize the Terman vocabulary. Willoughby (50), in a study of parents and children, found little decline with increase in age in vocabulary score when compared with other tests.

Jones and Conrad (24) gave the Army Alpha to 1,191 subjects ranging from age 10 through 60 and established that the Synonym-Antonym (vocabu-

lary) subtest did not show the post-adolescent decline found in all the tests (except the General Information subtest).

Sorenson (39), analyzing results from a group of 641 university extension students ranging up to age 69, reported increasing vocabulary scores for successively older age groups.

Christian and Paterson (9) gave Subtest *A* (vocabulary) of the University of Minnesota College Ability Tests to 329 University of Minnesota freshmen and their relatives. They discovered a steady increase in vocabulary up to the seventies, but pointed out that they were dealing with a superior population and that similar results might not hold for average or below average groups.

Wechsler (48) found that the Vocabulary Test of the Wechsler-Bellevue Scale, even though it did not remain constant right up to senescence, held up very well with age.

Thorndike and Gallup (45), administering an untimed, multiple-choice vocabulary test made up of items taken from the *IER* Intelligence Scale *CAVD* to a representative sample of the American adult voting public, reported that age differences for the test were almost nil, and that it was not until the "over sixty" age group was reached that any substantial drop in vocabulary score was discovered.

Sward (40), matching a superior population of 45 university professors aged 60 to 80 against a control group of 45 younger professors and instructors aged 25 to 35, concluded that in word knowledge or general vocabulary, the senescent are uniformly superior to the younger middle-aged.

Fox (14) matched a group of 30 subjects, aged 70 to 79, against an equivalent group of 30 subjects, aged 40 to 49, using both definition and multiple-choice types of vocabulary tests, and noted no difference in the vocabulary scores of the older group. She also indicated that people, regardless of age, found it more difficult to state the meaning of a word than to choose a synonym from among four alternatives.

The experimental evidence just cited ostensibly substantiates the contention that vocabulary, compared with other test functions, declines little, if at all, with increasing age. However, it is quite possible that these results may be explained by the crude scoring system utilized which tends to obscure characteristic differences in the quality of response in the definition type of vocabulary, and by the fact that in the multiple-choice type of vocabulary the correct answer is suggested even though the meaning of the word may have been forgotten. Since the qualitative aspects of intellectual performance have been acknowledged as being fully as important as its quan-

titative variations (Bolles, 5; Goldstein, 17; etc.), it appears that in our present state of knowledge, emphasis could be fruitfully placed on a qualitative analysis of vocabulary.

4. *Qualitative Development*

As far back as 1904 a very significant study of the growth in the general meaning of vocabulary words was made by Chambers (8). Testing 2,922 children and young people ranging in age from five to 27 years, he asked them to define and tell what they meant by the words "monk, peasant, emperor, nation, armor, and school." His analysis of the answers showed that the definitions could be classified under four headings: (*a*) No answer, signifying absence of content; (*b*) Wholly wrong answer; (*c*) Vaguely right answer, i.e., having one or more correct features; (*d*) Correct answer. The most common source of error he found to be "euphonic analogy" which was based on the similarity in sound of the response given to the particular word presented, e.g., "monkey" in response to the word "monk." He indicated that in the early years of life we have an accurate knowledge of only those things which are most immediate and familiar and that in studying the development of the child's use of language the most fruitful field is found in an analysis of the quality and expansion of vocabulary during the years of school life. He also drew attention to the fact that a correct definition in the early years is "a mere outline, a framework of bare essentials which in later years is filled in with various details" (8, p. 37).

About the same time Kirkpatrick indicated that growth with age occurred not only in the vocabulary range, but also in the character of the definitions attached to words. In this connection he reached the following conclusions as a result of analytical studies of definitions given by subjects of different ages:

> Descriptions which are so common in the high school and college papers are rarely or never given by children in the primary grades. The same is true of definitions by synonyms and inclusions under large terms. The younger children always define by mention of some specific incident, e.g., "a chair is to sit on," "baby stands up in a chair," "a bee goes around a piazza and makes a noise." What anything can do, or what can be done to it, or with it, is of most importance in early knowledge of all things, hence we find the definitions of children expressing action and use more than anything else (27, p. 17).

Binet and Simon (4), studying the development of intelligence in children, stated that children of age six define words in terms of "use" (i.e., what people do with it, what it does, etc.) and that it was not until they

reached age nine that most of the definitions they gave were "superior to use."

As early as the tentative revision of the Binet-Simon scale in 1912 Terman (41) considered a qualitative analysis of the vocabulary test important and stated that a qualitative analysis of the verbatim definitions of children of different ages and mentality would be of great value. In 1916 Terman (42) found that children of six, as a rule, defined objects in terms of use. Definitions utilizing "description" (i.e., telling what a substance is made of, etc.) and giving the "class" to which it belongs, were grouped together as superior to use. He reported that it was not before eight years that two-thirds of the children spontaneously gave a large proportion of definitions in terms superior to use. He also stated that the type of definition, i.e., whether thought of in terms of use, appearance, material comprising it, or class relationship, threw interesting light on the maturity of the child's apperceptive processes, and that it was "possible to differentiate, at least, a *half-dozen degrees of excellence in definition according to the intellectual maturity of the subject*" (42, *p.* 169). He admitted that although the form of the definition is significant it was not taken into consideration in scoring the test.

Dolch (12), studying reading and word meanings, found that among the types of words causing difficulty for children were those that acted as symbols for abstractions or generalizations, e.g., "advantage."

Marx (29), carrying out a qualitative study of the first 50 words of the 1916 Stanford-Binet vocabulary on a fairly large group of children and adults, reported that the highest quality types of definitions in relation to chronological age were those of the synonym and genus variety. Lowest on the scale were those definitions using illustration or example, use, and the repetition type of response.

Green (21) also qualitatively analyzed the responses of 718 school children and 110 adults on 50 vocabulary words (45 of which later became the Form *L* Vocabulary Test of the 1937 Revised Stanford-Binet). She worked out a method of weighting scores for each word in accordance with the relation between the quality of response and the developmental level of the subject. Although she used a somewhat different qualitative classification than Marx, her results followed the same trend. The data of Green's study led her to be dissatisfied with Binet's division of definitions into "use" and "superior to use." The implications of his method, she felt, were that all definitions other than in terms of "use" were of a type "superior to use." As sub-classifications, however, in this class of "superior to use" Binet considered (*a*) Synonyms, (*b*) Synonyms with modification by use, etc., (*c*)

Descriptions, (*d*) Grammatical explanations. She noted that ages six and seven were characterized by "use" definitions. As a result of what she considered the inadequacies of Binet's system of qualitative classification for word definitions and as a result of her investigations, Green devised an excellent qualitative classification system based on the type of thought of the child and adult.

She discovered that types of definition having lower median ages than "use" were "repetition in context, and demonstration." These classifications were included in Binet's group of incorrect responses. But "repetition" the way he conceived it was "repetition" pure and simple, e.g., "gown—it's a gown," whereas Green considered "repetition" to be a response that possessed some modification which convinced the tester that the child knew the meaning of the word, e.g., "gown—it's a nightgown." "Demonstration" she stated when used as a response to "eyelash" was more intelligent than when used as a response to "table" (example used by Binet). The most important part of a definition for the word "eyelash" is that it is the hair in the eyelid and not the eyebrow. The fact that the child pointed to his eyelash, she felt, was proof that he recognized this distinction.

She also noticed that young children perceived words as concrete ideas. The power of generalization has not yet fully developed sufficiently to make it possible for them to define "orange" as a "fruit," but rather as "you eat it." This bears out Piaget's (31) conclusion that the child can only reason about isolated or about more or less special cases. He cannot generalize from the particular. In addition, Green found that the "unmodified synonym" was most frequently employed by older children and adults to express the meaning of a word. But not always did the "unmodified synonym" express the exact shade of meaning of a word. In these cases the superior adult qualified the meaning by some limitation, such as "light touch" for "tap."

Gray and Holmes (20) emphasized that the character of the definitions attached to words changed notably from the lower to higher grades, and that the period from nine to 15 years was particularly productive in the acquisition of different types of meaning.

Watts discovered that the words and quality of the definition, in a well-developed language, could be arranged in a hierarchy of ascending levels of increasing generality, with those serving "as labels for concrete particular things at the lowest level, and those having reference to the most universal and abstract of concepts at the highest" (47, p. 53). He cited the interesting observation that the language of the more primitive peoples was usually lacking in general terms, e.g., the Eskimos have names for stones of various kinds, but no general word for all stones.

Wechsler (48) stated that in defining a word a subject gave much more than just its mere meaning. From the clinical point of view the character and quality of the word definition very often gave insight into the individual's thought processes.

Fox (14), in the study mentioned before, using a rough scoring system of one-half credits, found no significant change in the qualitative responses of her 70-year-old group as contrasted with her 40-year-old group.

Feifel and Lorge (13) qualitatively analyzed the verbatim responses of 900 children, between the ages of six and 14, to the Form *L* Stanford-Binet Vocabulary Test. Employing a five-fold qualitative category system, they discovered that the younger children significantly more often gave use and description types of definition, along with illustration, inferior explanation, demonstration, and repetition types of answer to the 45 words of the Test, whereas the older children significantly more often selected synonym types of definition.

Analysis of the qualitative studies of vocabulary emphasizes the fact that growth with age occurs not only in the vocabulary range but in the character of the word definition as well, and that the quality and completeness of the definition attached to words changes considerably from the lower to the higher grades. It indicates that there are many possible ways of defining a word, some of which are difficult, others comparatively easy. Both a six-year-old and a 14-year-old know the meaning of the word "orange." The six year-old defines it as "it's round," the 14-year-old as "a citrus fruit that grows in Florida." An older adult who was once capable of defining the word "gown" as "an evening dress worn to a party" now may be only capable of responding with "to put on." Obviously, in both instances, there are characteristic differences in thinking involved, and yet, in both cases, our present scoring of the vocabulary test gives similar credit to the responses.

5. *Mental Impairment*

The use of a coarse "all or none" system for scoring the vocabulary test, with its tendency to obscure the quality and precision of the word definition and consequently possible differences in the thinking processes, may help throw some light on another prominent generalization from the literature on intellectual change during the past 20 years. This has been the alleged stability of vocabulary test scores in cases of the mentally impaired and psychotics. This accepted premise together with the assumption that vocabulary ability does not decline as a function of age has led to a utilization of the vocabulary test by many in mental testing as an indicator of the original level of endowment in abnormals.

Along this line, Babcock (3) developed her Mental Efficiency Test. It was based on the assumption that mental deterioration occurred first in new learning and in the formation of new associations, and last in earliest formed material (vocabulary). The 1916 Terman vocabulary was employed as the measure of original mental level and a series of 30 speed and new learning tests was used for getting the efficiency phase of intelligence. Each of the latter tests was scored in terms of mental level. The discrepancy between the scores on the two types of tests was termed "the efficiency index." This discrepancy was close to zero in normal individuals and increased with the severity of the condition in the psychotics.

Schwartz (36) followed Babcock's technique in testing 110 dementia praecox patients and confirmed her general belief in the preservation of vocabulary in psychosis. He found that the language ability of the deteriorated patient was most resistive to the ravages of the disturbance, and that the difference between the vocabulary and non-vocabulary test indicated the degree of mental impairment present.

Their evidence concerning vocabulary constancy in psychosis was more or less supported by the work of Wittman (51), Jastak (23), Davidson (10, 11), Malamud and Palmer (28), etc. Shipley (38), using a self-administering scale, indicated that the discrepancy between scores on vocabulary and on a 10-minute test of abstract thinking could be used as a sensitive measure of intellectual impairment.

Nevertheless, the Babcock technique has been criticized on two grounds. Theoretical objection has been raised by Yacorzynski (52, 53) who said that the reason the Babcock test appeared valid was not because recently acquired behavior (therefore easily lost) was compared to old habits (vocabulary, and therefore well maintained), but because "goal activity which can be carried to completion with only one set of acts is compared with goal activity where end-results can be achieved by a number of separate and qualitatively different acts of unequal difficulty" (52, p. 425). In other words, most of the items in the Babcock test gave the subject only one or at best a few methods of solving the problem, whereas on the vocabulary test, the subject had a number of choices with which to define the word, and if even the more difficult conceptual organization had deteriorated, the correct response could still be given on a "lower" conceptual level. To test his hypothesis Yacorzynski devised a one-alley maze which could be solved by seven appropriate but different methods. Employing human subjects, he showed that there was a qualitative difference in the methods used to solve the problem related to the general level of mental ability of the subjects. Goldstein (17) presented

some pertinent data on this question. He pointed out that certain brain lesions did not destroy the ability to solve a problem but affected the method by which the problem could be solved. Rapaport (34), as a result of his work with the mentally impaired, felt that the main reason for the stability of the vocabulary test in intelligence testing was because responses on different levels of abstraction were acceptable.

The Babcock technique has also been criticized on the ground that vocabulary ability by no means always remains unimpaired in psychosis. Capps (7) demonstrated a severe loss in epileptics with consequent invalidation of the Babcock index. He tested 80 idiopathic epileptics and 20 normals aged 16 to 50 and found that the 1916 Terman vocabulary along with other vocabulary tests did not remain constant and unchanged with increasing deterioration in his epileptic patients, and therefore concluded that they could not be employed to determine the native intellectual capacity of his patients.

Kendig and Richmond (26) had reservations regarding the use of the 1916 Terman vocabulary as a criterion of prepsychotic mental level as a consequence of their work with 500 dementia praecox cases. They felt that the Babcock Efficiency Index type of approach might have relative but not absolute value, and that vocabulary level was better preserved than other mental functions, but by no means a satisfactory measure of the original endowment level.

Gottschalk (19), in his work with epileptics, believed that the Babcock deterioration scale was not a reliable quantitative measure of deterioration in any but mild cases. He discovered that as deterioration increased the reliability of the index decreased.

Roe and Shakow (35), using the 1916 Terman vocabulary in a study of 827 mental patients, implied that in severe psychosis the vocabulary score was a better indicator than the total Stanford-Binet score, but clearly indicated that it was, nevertheless, far from an adequate index of original endowment since it was far below the level of previous education in their paretic and hebephrenic groups.

Rapaport (34), employing the Bellevue-Wechsler Scale with neurasthenics, psychotic depressives, and deteriorated schizophrenics, discovered that the Vocabulary subtest while holding up much better than did the other subtests, nevertheless, showed a strong decline from its previous level.

Acklesberg (1), working with senile dementia patients, noted that vocabulary functioning as measured by her tests (similar to the ones used by Capps, e.g., synonyms, antonyms, word-naming, homographs, word categorizations, etc.), did not remain constant or unchanged in the type of deterioration present in senile dementia.

It is manifest that in advanced organic cases, deteriorated schizophrenics, and in certain types of epilepsy the vocabulary score does decline along with other mental functions. Yacorzynski (53) pointed out that there is little reason to suppose that even if only the end results of the ability to define words are measured, the vocabulary scores of mentally impaired patients will remain unaffected, since even the easiest definitions of some words are too difficult for the lower mental age groups. Capps (7), as a matter of fact, drew attention to the direct relationship between the degree of deterioration in his idiopathic epileptics and their scores on certain tasks which essentially measured the ability to use words. He reported that scores on written tests (except for antonyms) correlated to a higher degree with deterioration than the scores of tests on which the subject responded orally. Examination of the kinds of tests which he used showed that the correct answers to the written tests were predetermined and allowed the subject no alternative in obtaining the correct solution, but that the oral tests (Stanford-Binet vocabulary, homographs, and free word associations) could be answered by many acceptable responses. This is the same condition, as Yacorzynski has stated, which obtains between the items of the Babcock test and the vocabulary score and can, therefore, be explained on the same basis, that is to say, if a number of methods are available for reaching an end result, the more difficult solutions need not be available to the subject in order to obtain the correct response.

B. Present Problem

The Yacorzynski hypothesis states that the reason the vocabulary of older adults does not decline as do other tests, and the reason the vocabulary scores of mentally impaired persons show a minimal amount of change from their pre-impaired level, may be because of the crude system used in scoring the vocabulary test and its consequent eclipsing of the fact that there are many appropriate solutions of varying difficulty possible in defining a word, and that the final result of defining a word can be made on the basis of easier definitions even if the mechanisms involved in the more difficult conceptual organization are no longer available. Verification of this supposition demands qualitative analysis of the definitions of words used by normal and abnormal subjects. Research in this area is needed because: (*a*) No work is evident in the literature on any qualitative analysis, as suggested above, of the vocabulary definitions of abnormal adults. (*b*) Neither has any comparison ever been made between the qualitative word definitions of similar groups of normal and abnormal adults. (*c*) Vocabulary score as an indicator of

the original level of the subject is quite extensively used in many clinics and mental institutions. Qualitative analysis of vocabulary definitions may present us with a satisfactory measure of the possible loss that vocabulary itself sustains in older adults and in abnormal patients. (*d*) Further information may be secured on developmental stages in conceptual thinking and aspects of mental functioning which cannot be revealed by techniques simply involving a pass or fail scoring system.

The present experiment provides: (*a*) A qualitative analysis of the vocabulary definitions of normal adults. (*b*) A qualitative analysis of the vocabulary definitions of psychotics. (*c*) A comparison between the qualitative vocabulary definitions of normal and abnormal groups matched on the factors of age, education, and score correct on the Stanford-Binet Vocabulary Test.

By means of a qualitative analysis of the verbatim responses given by normal and abnormal subjects, ranging in age from 15 to 80 years of age, to the 45 words of the Form *L* Stanford-Binet Vocabulary Test, this investigation attempts to answer the following questions: (*a*) Is there a qualitative vocabulary decline in our older normal population? (*b*) Is there a qualitative vocabulary decline in our older psychotic population? (*c*) Can we differentiate between our normal and abnormal population?

II. PROCEDURE OF INVESTIGATION

A. Selection of the Test

The present study utilized the Form *L* Vocabulary Test of the 1937 Revised Stanford-Binet Tests of Intelligence which consists of 45 words, graded in difficulty, to which an acceptable definition must be given in order to receive credit. It was chosen because it extends from the six-year through the superior adult level, and the definitions given to its 45 words allow for quality differences in the responses to express themselves. In addition, the Test has excellent standardization, good interest value, and, generally, is more desirable for use with abnormals than the vocabulary paper and pencil type of test.

However, one should bear in mind that not all of the words of the Test permit a full range of qualitative differences to appear in the verbatim responses. This is especially true for the more difficult words in the list. Whereas for a word like "envelope" one can reply with various types of acceptable answers, e.g., "container" (synonym), "receptacle for paper" (synonym modified), "you mail it" (use), "white paper" (description), etc., the correct answers for a word like "homunculus" are usually limited to a synonym or synonym type of definition, e.g., "some sort of variety of dwarf" or "little man."

B. Qualitative Classification of Responses

On the basis of Green's (21) well developed system for classifying the different definitions of words, and on an analysis of verbatim responses in a preliminary investigation by the author, it was discovered that all the variety of different qualitative responses given to all the words could be encompassed in five essential or key categories. With this information a five-fold qualitative classification system was set up for each definition of every word. One category consisted of synonym types of response. Another included use and description types of response. These were combined into one category because they were found to occur at approximately the same developmental level in children. A third category contained the explanation type of response. A fourth one was made up of the illustration, demonstration, inferior explanation, and repetition types of response. These were included together because empirical analysis also indicated that they occurred at about the same developmental level. The final category was composed of all types of error response. These categories were kept uniform for the entire list of 45 words. Here are examples for each of the five categories:

Synonym Category

(*a*)	Synonym unmodified:	Orange = a fruit.
(*b*)	Synonym modified by use:	Straw = hay that cattle eat.
(*c*)	Synonym modified by description:	Gown = long dress.
(*d*)	Synonym modified by use and description:	Eyelash = hair over the eye that protects you.
(*e*)	Synonym qualified as to degree:	Tap = touch lightly.

Use, Description, and Use and Description Category

(*a*)	Use:	Orange = you eat it.
(*b*)	Description:	Straw = it's yellow.
(*c*)	Use and Description:	Orange = you eat it and it's round.

Explanation Category

(*a*)	Explanation:	Priceless = it's worth a lot of money. Skill = being able to do something well.

Illustration, Inferior Explanation, Repetition, and Demonstration Category

(*a*)	Illustration:	Priceless = a gem.
(*b*)	Inferior Explanation:	Scotch = hot.
(*c*)	Repetition:	Puddle = puddle of water.
(*d*)	Demonstration	For words like tap, eyelash, etc.

Error Category

(Incorrect Demonstration, Misinterpretation, Wrong Definition, Clang Association, Repetition Without Explanation, Omits)

(*a*)	Incorrect Demonstration:	Eyelash = points to eyebrow.
(*b*)	Misinterpretation:	Regard = protects something.
(*c*)	Wrong Definition:	Orange = a vegetable.
(*d*)	Clang Association:	Roar = raw; Skill = skillet.
(*e*)	Repetition Without Explanation:	Puddle = puddle.
(*f*)	Omit:	When the word is left out.

In all cases used in this study the Vocabulary Test was first administered and scored according to Terman's (44) directions by experienced examiners. After checking the quantitative scoring the author rescored all of the verbatim word definitions in terms of the five qualitative categories just outlined.

C. Reliability of Scoring

In measuring the reliability of qualitative categories a common procedure is to compare, item by item, or response by response, the records obtained when two or more independent workers record the same behavior, or score the same responses. Agreement can be computed in terms of percentages. To establish the reliability of the qualitative scoring, 54 of the cases were rescored by an experienced examiner not specially trained in the qualitative method of scoring. Employing the Arrington (*2*) formula where the responses in *each* observer's scoring that agree with the other's (in effect, doubling the agreements) is divided by this total plus the disagreements (responses dissimilarly recorded, and responses noted by one observer and omitted by the other),[2] i.e., $\frac{2 \times \text{agreements}}{2 \times \text{agreements} + \text{disagreements}}$, the range of the per cent of agreements for the 45 words of the Test was found to extend from 96 to 100 per cent. They are of such an order to signify that the qualitative scoring was carried out with a high degree of consistency.

D. Subjects

The tested population consisted of 185 normal and 185 abnormal adults, male and female, ranging in age from 15 to 80 years. Data for the normal subjects were secured from testing done by the author and from testing carried out in two advanced psychological testing courses at Teachers College, Columbia University. All these individuals were tested in 1947. The abnormal population data were obtained by the author from the extensive records of St. Elizabeths Hospital in Washington, D. C. Previous inquiry had revealed that it was practically the only institution that had verbatim responses of the 1937 Stanford-Binet Vocabulary Test available on a large sampling of adult abnormal subjects. The 185 abnormal patients used in the study were selected from a group of 476 cases which were originally secured at the Hospital. There is no reason to believe that the 185 cases chosen for the study were significantly different in any manner from those not used. Of the original 476 cases obtained approximately five or six had to be discarded because they did not have complete verbatim responses for every word. The sample consisted essentially of schizophrenics, manic-depressives, and organics, with schizophrenics predominating. Since testing is given

[2] There were no responses in any of the 54 cases noted by one observer and not by the other. *This part applies essentially to workers recording observational types of* behavior.

shortly after the patient enters the hospital and before he appears at an admission conference for preliminary diagnosis, all of the subjects used were tested relatively soon after the recognition of the psychosis and the resort to hospitalization. None of the cases was tested more than three months after entry into the institution. It should be remembered that since some of the patients at St. Elizabeths Hospital are transfers from other mental institutions testing, in those cases, would actually have occurred more than three months after recognition of the disorder. However, in the sample used for the study only a negligible minority were in that category. These individuals were tested during the years 1940-1947. Testing was carried out by trained examiners with considerable experience in examining psychotic subjects. Both the normal and abnormal populations in this study came mainly from the Eastern seaboard states with a scattering of Midwest states also represented.

All subjects were of the white race, American born (except for a very few cases where, however, the subject's education had been received in this country), and gave no indication of any language or physical handicaps. All the test data were representative of the individual's abilities. This was based on the subject's test behavior and attitude as well as test performance, and the judgment of the examiners. Whenever any question arose as to the validity of the data involved, the case was discarded.

In the first tabulations, results for males and females were kept separate. Analysis showed no significant sex differences but rather very similar pictures. These findings agree with those of Capps (7) and Roe and Shakow (35). As a consequence the data for the sexes have been combined and only total group determinations, in this respect, are reported in the present study.

In determining educational level, this study follows earlier investigations and credits either completion of or some attendance at a grade in establishing the number of years of schooling.

The subjects were divided into a principal group of 135 normals and 135 abnormals matched on the variables of age, education, and score correct on the Stanford-Binet Form *L* Vocabulary Test. Age was subdivided into three divisions: 15-29, 30-49, and 50-79.

Table 1 presents the relevant information concerning the background of this group (Group 1). It indicates that the group of 135 normals and 135 abnormals is very well matched on the factors of age, education, and score correct on the Stanford-Binet Vocabulary Test. Since abnormals, in general, tend to have a lower educational level than normals of presumable

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR AGE, EDUCATION, AND SCORE CORRECT OF GROUP I BY AGE CLASSIFICATIONS (N = 270)*

Age classifications		Age		Highest grade reached		Score correct		
		Mean	*SD*	Mean	*SD*	Mean	*SD*	*N*
15-29	Normal	21.9	3.5	11.8	1.9	24.1	4.3	38
	Abnormal	22.5	3.3	11.9	1.9	24.0	4.5	38
30-49	Normal	39.2	5.4	10.7	2.9	24.3	6.1	56
	Abnormal	38.9	5.2	10.7	2.9	24.4	6.1	56
50-79	Normal	59.0	7.1	10.1	2.6	24.1	6.2	41
	Abnormal	58.8	7.0	10.1	2.6	24.1	6.2	41

*Group I = 135 normals matched against 135 abnormals on the variables of age, education, and score correct.

similar intelligence, it is possible that the abnormals of this group may be superior to the normals against whom they are matched.

Another supplementary group of 50 normals and 50 abnormals was matched on the factors of age and score correct on the Form *L* Vocabulary Test to see what differences would appear when the education factor was uncontrolled. Here, too, age was subdivided into three main divisions: 15-29, 30-49, and 50-79.

Table 2 presents the background of this group (Group II). It shows that this group of 50 normals and 50 abnormals is well matched on the variables

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR AGE, EDUCATION, AND SCORE CORRECT OF GROUP II BY AGE CLASSIFICATIONS (N = 100)

Age classifications		Age		Highest grade reached		Score correct		
		Mean	*SD*	Mean	*SD*	Mean	*SD*	*N*
15-29	Normal	21.4	3.4	13.1	1.7	23.5	2.9	30
	Abnormal	21.5	3.8	8.5	1.6	23.2	2.8	30
30-49	Normal	35.5	4.7	14.4	1.8	24.8	4.1	16
	Abnormal	36.2	4.3	7.7	3.2	24.6	4.1	16
50-79	Normal	59.5	6.5	8.5	1.8	18.8	6.1	4
	Abnormal	58.5	5.4	3.0	1.6	18.8	6.4	4

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR AGE, EDUCATION, AND SCORE CORRECT OF GROUP I AND II COMBINED BY AGE CLASSIFICATIONS (N = 370)*

Age classifications		Age		Highest grade reached		Score correct		
		Mean	*SD*	Mean	*SD*	Mean	*SD*	*N*
15-29	Normal	21.6	3.5	12.4	1.9	23.8	3.7	68
	Abnormal	22.0	3.5	10.4	2.4	23.6	3.9	68
30-49	Normal	38.3	5.5	11.5	3.1	24.4	5.7	72
	Abnormal	38.3	5.0	10.0	3.2	24.5	5.7	72
50-79	Normal	59.0	7.1	9.9	2.5	23.7	6.4	45
	Abnormal	58.8	6.8	9.4	3.6	23.7	6.4	45

*Group II = 50 normals matched against 50 abnormals on the variables of age and score correct.

of age and score correct on the Form L Vocabulary Test, and that there is quite a disparity in the educational background of the normals and abnormals.

The characteristics of both Group I and Group II combined are summarized in Table 3.

The data in Table 3 indicate that the normals and abnormals of Groups I and II combined are very well matched on the factors of age and score correct on the Vocabulary Test, and to a lesser degree on educational attainment.

According to Terman (44) the score of the average adult on the Form L Stanford-Binet Vocabulary Test is about 20 words correct. It appears, therefore, that both Group I and Group II are above average in mental ability.

III. TREATMENT OF THE DATA

The necessary data for answering the problem of this investigation are presented in the succeeding tables. The basic statistical treatment used was the critical ratio technique. In common with other statistical measures of significance, the critical ratio furnishes a measure of the validity of the null hypothesis, i.e., a measure of the probability that an obtained result may be attributable to chance. A critical ratio between 1.96 and 2.57 is considered as questioning the validity of the null hypothesis and the finding is viewed as significant. A critical ratio of 2.58 and higher seriously questions the validity of the null hypothesis, and the finding is regarded as highly significant. The critical ratio is represented by the ratio of the obtained difference to the standard error of the difference.

A. Differences in Qualitative Responses between Normals and Abnormals

In order to determine whether there were any significant differences between the type of qualitative definitions given by the normals as against the abnormals to the 45 words of the Stanford-Binet Vocabulary Test, the mean differences were obtained for both groups of subjects and these tested for significance. Table 4 gives the results for Group I composed of 135 normals and 135 abnormals matched on age, education, and score correct on the Form *L* Vocabulary Test, for Group II comprising 50 normals and 50 abnormals matched on age and score correct on the Vocabulary Test, and for Groups I and II combined. They are given separately for the age ranges 15-29, 30-49, and 50-79, and then combined for all age ranges.

It is evident from Table 4 that, at all age ranges, the normals of Group I, Group II, and Groups I and II combined choose the synonym types of response significantly more often than do the abnormals. In the same manner, the abnormals in all the Groups select significantly more often than do the normals the use and description types of response, and the demonstration, illustration, repetition, and inferior explanation types of answer. The abnormals also pick the explanation type of response significantly more often than do the normals at all age ranges except for the 15-29 age classification of Group I, and the 50-79 age classification of Group II.

No significant differences were found between the normal and abnormal subjects on the error category for any of the Groups at any of the age classifications. This is to be expected since the normals and abnormals of both Group I and Group II were originally matched on total score correct on the Vocabulary Test.

TABLE 4

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS AND ABNORMALS OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES BY AGE CLASSIFICATIONS

	Group I $N = 58$				Group II $N = 30$				Groups I and II $N = 68$			
Age 15-29	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_n	5.57	.81	1.67	.71	3.91	.88	1.89	.88	4.91	.85	1.78	.79
SD_a	5.79	1.48	2.51	1.32	3.75	1.14	1.87	.89	5.01	1.35	2.26	1.18
Mean_n	18.95	.60	4.80	.50	18.54	.77	5.56	.63	18.76	.68	3.81	.56
Mean_a	15.79	1.71	4.76	1.71	14.87	1.77	5.43	1.16	15.38	1.74	5.06	1.47
$\overline{D}$†	3.16	—1.11	— .76	—1.21	3.67	—1.00	—1.87	— .55	3.38	—1.06	—1.25	— .91
$SE_{\overline{d}}$‡	.73	.21	.53	.26	.62	.25	.47	.19	.49	.16	.37	.17
$\overline{D}/SE_{\overline{d}}$	4.33**	—5.28**	—1.43	—4.65**	5.92**	—4.00**	—3.98**	—2.79**	6.80**	—6.63**	—3.38**	—5.35**

	Group I $N = 56$				Group II $N = 16$				Groups I and II $N = 72$			
Age 30-49	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_n	6.96	.87	2.15	1.04	4.80	.91	2.49	.68	6.53	.89	2.25	.98
SD_a	6.80	1.31	2.80	1.47	6.17	1.10	2.43	1.61	6.67	1.27	2.75	1.51
Mean_n	18.21	.77	4.44	.91	19.25	.75	4.06	.69	18.46	.76	4.36	.86
Mean_a	14.98	1.88	5.78	1.80	14.25	1.65	6.62	2.13	14.83	1.82	5.97	1.87
$\overline{D}$	3.23	—1.11	—1.34	— .89	5.00	— .88	—2.56	—1.44	3.63	—1.06	—1.61	—1.01
$SE_{\overline{d}}$	.60	.20	.48	.24	.95	.32	.80	.33	.51	.17	.41	.20
$\overline{D}/SE_{\overline{d}}$	5.38**	—5.56**	—2.79**	—3.71**	5.26**	—2.75**	—3.20**	—4.36**	7.12**	—6.24**	—3.93**	—5.05**

†Difference between obtained means for normal and abnormal subjects.
—Minus values indicate that the difference is in favor of the abnormals.
**Significant at .01 level.

‡$SE_{\overline{d}} = \dfrac{SD_d}{\sqrt{N}}$.

TABLE 4 (*continued*)

Age 50-79	Group I $N = 41$				Group II $N = 4$				Groups I and II $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_s	7.45	1.08	2.79	.96	6.18	.71	1.50	.00	7.61	1.05	2.74	.94
SD_d	6.94	1.28	2.52	1.55	4.55	1.41	3.19	.83	7.00	1.33	2.48	1.51
$Mean_s$	17.71	1.10	4.61	.73	13.25	1.00	4.50	.00	17.14	1.08	4.55	.65
$Mean_d$	13.32	2.25	5.88	2.71	8.25	3.00	5.25	2.25	12.74	2.29	5.76	2.65
$\bar{D}$†	+.39	—1.15	—1.27	—1.98	5.00	—2.00	— .75	—2.25	+.40	—1.21	—1.21	—2.00
$SE_{\bar{d}}$‡	.60	.25	.53	.25	1.30	.91	1.38	.48	.56	.24	.49	.25
$\bar{D}/SE_{\bar{d}}$	7.32**	—4.60**	—2.40*	—7.92**	3.85**	—2.20*	— .54	—4.69**	7.86**	—5.04**	—2.47*	—8.70**

Age 15-79	Group I $N = 135$				Group II $N = 50$				Groups I and II $N = 185$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_s	6.77	.95	2.27	.95	4.68	.88	2.10	.80	6.30	.95	2.25	.92
SD_d	6.66	1.37	2.62	1.52	5.03	1.21	2.26	1.27	6.27	1.33	2.54	1.46
$Mean_s$	18.27	.82	4.37	.74	18.34	.78	3.80	.60	18.27	.81	4.21	.70
$Mean_d$	14.71	1.94	5.53	2.05	14.14	1.82	5.80	1.56	14.54	1.91	5.59	1.92
$\bar{D}$	3.56	—1.12	—1.16	—1.31	4.20	—1.04	—2.00	— .96	3.73	—1.10	—1.38	—1.22
$SE_{\bar{d}}$	.37	.13	.30	.15	.49	.19	.39	.18	.30	.11	.24	.12
$\bar{D}/SE_{\bar{d}}$	9.62**	—8.62**	—3.87**	—8.73**	8.57**	—5.47**	—5.13**	—5.33**	12.43**	—10.00**	—5.75**	—10.17**

†Difference between obtained means for normal and abnormal subjects.
—Minus values indicate that the difference is in favor of the abnormals.
*Significant at .05 level.
**Significant at .01 level.

‡$SE_{\bar{d}} = \dfrac{SD_d}{\sqrt{N}}$.

B. Differences in Qualitative Responses between Young and Old Normals

To answer the question of whether any important differences exist in the quality of the definitions given to the 45 words of the Form *L* Vocabulary Test by the old as compared with the young normals, the differences in the mean category scores for the various age levels, 15-29, 30-49, and 50-79, were determined, and an evaluation made of the significance of these differences in terms of the standard error of their differences.

The results for the normals of Group I and Group II, and Groups I and II combined appear in Tables 5, 6, and 7.

The findings in Tables 5, 6, and 7 indicate that for Group I there do not seem to be any consistent reliable statistical differences between the younger and older normals. However, the 30-49 age classification does pick more often the demonstration, repetition, inferior explanation, and illustration types of response than does the 15-29 age classification. The results for Group II show that the only important difference in the types of definition given by the younger as against the older normal subjects is that the 15-29 and 30-49 age classifications choose the demonstration, repetition, inferior explanation, and illustration kinds of response significantly more often than does the 50-79 age classification. This appears in contradistinction to the results for Group I where the old rather than the young normals tend to select the inferior explanation, illustration, demonstration, and repetition types of response. This seeming difference may be explained by the small number of cases used in the 50-79 age classification. The data for Groups I and II combined indicate that some significant differences do exist, at the .05 level, between the types of response chosen by the older as compared with the younger normals. The 30-49 age group selects more often the inferior explanation, demonstration, illustration, and repetition types of answer than does the 15-29 age group, and the 50-79 age group picks more often the use and description kinds of definition than does the 15-29 age group.

Although the older normals tend to select more often the use and description kinds of response than do the younger normals, on the whole, no clear differences appear to exist between the younger and older normals in their choice of types of response. It should be borne in mind that the categories are interdependent and that a significant choice of one kind of category will necessarily affect the amount of choice for the others.

TABLE 5

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS (15-29) AND NORMALS (30-49) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES

	Group I *a* 15-29; $N = 38$ *b* 50-49; $N = 56$				Group II *a* 15-29; $N = 50$ *b* 30-49; $N = 16$				Groups I and II *a* 15-29; $N = 68$ *b* 30-49; $N = 72$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	5.57	.81	1.67	.71	5.91	.88	1.89	.88	4.91	.85	1.78	.79
SD_b	6.96	.87	2.15	1.04	4.80	.91	2.49	.68	6.55	.89	2.25	.98
Mean_a	18.95	.60	4.00	.50	18.54	.77	3.56	.63	18.76	.68	3.81	.56
Mean_b	18.21	.77	4.44	.91	19.25	.75	4.06	.69	18.46	.76	4.36	.86
$\bar{D}_{a-b}$	.74	— .17	— .44	— .41	— .71	.02	— .50	— .06	.30	— .08	— .55	— .30
$\tilde{\sigma}_{a-b}$	1.30	.17	.39	.17	1.43	.28	.73	.24	.97	.14	.55	.14
$\frac{\bar{D}_{a-b}}{\tilde{\sigma}_{a-b}}$	.57	—1.00	—1.13	—2.41*	— .50	.07	— .68	— .25	.31	— .57	—1.57	—2.14*

$$\tilde{\sigma}_{a-b} = \sqrt{\frac{\sigma^2 15\text{-}29}{N_1} + \frac{\sigma^2 30\text{-}49}{N_2}}.$$

—Minus values indicate that the difference is in favor of older group.

*Significant at .05 level.

TABLE 6

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS (15-29) AND NORMALS (50-79) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES

	Group I *a* 15-29; $N = 38$ *c* 50-79; $N = 41$				Group II *a* 15-29; $N = 30$ *c* 50-79; $N = 4$				Groups I and II *a* 15-29; $N = 68$ *c* 50-79; $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	5.57	.81	1.67	.71	3.91	.88	1.89	.88	4.91	.85	1.78	.79
SD_c	7.45	1.08	2.79	.96	6.18	.71	1.50	.00	7.61	1.05	2.74	.94
$Mean_a$	18.95	.60	4.00	.50	18.54	.77	3.56	.65	18.76	.68	3.81	.56
$Mean_c$	17.71	1.10	4.61	.73	13.25	1.00	4.50	.00	17.14	1.08	4.55	.65
$\bar{D}_{a-c}$	1.24	— .50	— .61	— .23	5.29	— .25	— .94	.63	1.62	— .40	— .74	— .09
$\tilde{\sigma}_{a-c}$	1.47	.22	.51	.17	3.64	.45	.93	.17	1.27	.17	.46	.17
$\frac{\bar{D}_{a-c}}{\tilde{\sigma}_{a-c}}$	.84	—2.27*	—1.20	—1.35	1.45	— .51	—1.01	3.71**	1.28	—2.35*	—1.61	— .53

—Minus values indicate that the difference is in favor of older group.
*Significant at .05 level.
**Significant at .01 level.

TABLE 7

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS (30-49) AND NORMALS (50-79) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES

	Group I *b* 30-49; $N = 56$ *c* 50-79; $N = 41$				Group II *b* 30-49; $N = 16$ *c* 50-79; $N = 4$				Groups I and II *b* 30-49; $N = 72$ *c* 50-79; $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_b	6.96	.87	2.15	1.04	4.80	.91	2.49	.68	6.55	.89	2.25	.98
SD_c	7.45	1.08	2.79	.96	6.18	.71	1.50	.00	7.61	1.05	2.74	.94
$Mean_b$	18.21	.77	4.44	.91	19.25	.75	4.06	.69	18.46	.76	4.36	.86
$Mean_c$	17.71	1.10	4.61	.73	13.25	1.00	4.50	.00	17.14	1.08	4.55	.65
$\bar{D}_{b-c}$	.50	— .33	— .17	.18	6.00	— .25	— .44	.69	1.32	— .32	— .19	.21
$\tilde{\sigma}_{\bar{D}_{b-c}}$	1.49	.20	.52	.20	3.78	.47	1.08	.17	1.36	.17	.48	.17
$\frac{\bar{D}_{b-c}}{\tilde{\sigma}_{\bar{D}_{b-c}}}$	.34	—1.65	— .33	.90	1.56	— .53	— .41	4.06**	.97	—1.88	— .40	1.24

—Minus values indicate that the difference is in favor of older group.
**Significant at .01 level.

C. Differences in Qualitative Responses between Young and Old Abnormals

To find an answer to the question of whether any important differences exist in the quality of the definitions given to the 45 words of the Form *L* Vocabulary Test by the young as compared with the old abnormals, analysis was made, as with the normals, of the differences in the mean category scores for the various age levels, and evaluation made of the significance of these differences in terms of the standard error of their differences.

The results for the abnormals of Group I, Group II, and Groups I and II combined appear in Tables 8, 9, and 10.

The findings in Tables 8, 9, and 10 show that for Group I the abnormals, aged 50-79, choose the inferior explanation, demonstration, illustration, and repetition types of definition significantly more often than do the abnormals, aged 15-29, and 30-49. The 50-79 age classification also picks more often the explanation answer type, at the .05 level, than does the 15-29 age classification. The data indicate that for Group II the abnormals aged 15-29 select the synonym types of response significantly more often than do the abnormals aged 50-79. The age groups 30-49, and 50-79 also pick the demonstration, inferior explanation, repetition, and illustration types of answer significantly more frequently than the 15-29 age group. This tendency for the older abnormals to prefer the inferior explanation, repetition, illustration, and demonstration category types of response was also found in Group I. The results for the abnormals of Groups I and II combined show that, at the .05 level, the younger abnormals, aged 15-29, choose significantly more often the synonym kinds of definition than do the older abnormals, aged 50-79. The older abnormals, aged 50-79, select significantly more often, at the .05 level, the use and description types of response than do the younger abnormals, aged 15-29, and 30-49. The 30-49 age group selects the explanation type of definition more often than does the 15-29 age group. With reference to the demonstration, repetition, illustration, and inferior explanation types of response the older abnormals, aged 50-79, significantly more often choose them, at the .01 level, than do the younger abnormals aged 15-29, and 30-49.

On the whole, the younger abnormals appear to prefer the genus species and synonym types of answer; the older abnormals the use and description, explanation, and demonstration, illustration, inferior explanation, and repetition types of answer. It should be recalled that the categories are interdependent and that significant choice of one kind of category will necessarily affect the amount of choice for the others.

TABLE 8

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN ABNORMALS (15-29) AND ABNORMALS (30-49) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES

	Group I *a* 15-29; $N = 38$ *b* 30-49; $N = 56$				Group II *a* 15-29; $N = 30$ *b* 30-49; $N = 16$				Groups I and II *a* 15-29; $N = 68$ *b* 30-49; $N = 72$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	5.79	1.48	2.51	1.32	3.73	1.14	1.87	.89	5.01	1.35	2.26	1.18
SD_b	6.80	1.31	2.80	1.47	6.17	1.10	2.45	1.61	6.67	1.27	2.75	1.51
$Mean_a$	15.79	1.71	4.76	1.71	14.87	1.77	5.43	1.16	15.38	1.74	5.06	1.47
$Mean_b$	14.98	1.88	5.78	1.80	14.25	1.65	6.62	2.13	14.83	1.82	5.97	1.57
$\bar{D}_{a-b}$	.81	— .17	—1.02	— .09	.62	.14	—1.19	— .97	.55	— .08	— .91	— .40
$\tilde{\sigma}_{a-b}$	1.31	.30	.56	.30	1.75	.35	.71	.45	.99	.22	.44	.22
$\frac{\bar{D}_{a-b}}{\tilde{\sigma}_{a-b}}$	.62	— .57	—1.82	— .30	.36	.40	—1.68	—2.16*	.56	— .36	—2.07*	—1.82

—Minus values indicate that the difference is in favor of older group.

*Significant at .05 level.

An analysis of the psychoses of the abnormal subjects used in this study was carried out to check the possibility that the results found between the younger and older abnormals might be due to possible differences in the types of psychoses at the different age classifications. No significant differences in the types of psychoses were discovered for either Group I, Group II, or Groups I and II combined between any of the age classifications. Table 11 gives a distribution of the types of psychoses for the three groups at the different age classifications.

TABLE 11

DISTRIBUTION OF TYPES OF PSYCHOSES OF ABNORMALS OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED BY AGE CLASSIFICATIONS

Age classification	Type of Psychosis	Group I		Group II		Groups I and II	
		N	%	*N*	%	*N*	%
15-29	Schizophrenics	24	64	20	66	44	64
	Organics	7	18	5	17	12	18
	Manic-Depressives	7	18	5	17	12	18
		38	100	30	100	68	100
30-49	Schizophrenics	37	66	8	50	45	63
	Organics	11	20	2	13	13	18
	Manic-Depressives	8	14	6	37	14	19
		56	100	16	100	72	100
50-79	Schizophrenics	22	53	2	50	24	54
	Organics	8	20	1	25	9	20
	Manic Depressives	11	27	1	25	12	26
		41	100	4	100	45	100

A comparison of the qualitative responses to the 45 words of the Vocabulary Test between the normal and abnormal subjects used in this study indicates the following: (*a*) The normals significantly more often choose the synonym types of response, whereas the abnormals significantly more often choose the use and description, explanation, and demonstration, repetition, inferior explanation, and illustration types of definition. (*b*) No clear differences appear between the younger and older normals in their selection of the synonym, explanation, and demonstration, inferior explanation, repetition, and illustration types of response. (*c*) The younger abnormals tend to pick more often the synonym types of answer than do the older abnormals. (*d*) The older normals and abnormals tend to select more often the use and description kinds of definition than do the younger normal and abnormal subjects. (*e*) No distinct differences appear between the younger and older

abnormals in their choice of the explanation type of response. (*f*) The older abnormals choose significantly more often the inferior explanation, demonstration, illustration, and repetition types of definition than do the younger abnormals.

D. Differences in Qualitative Responses on First 10 Words of Stanford-Binet Vocabulary Test

Although one of the factors on which both Groups I and II were matched was the score correct received on the 45 words of the Stanford-Binet Vocabulary Test, it is evident that the total score received can be determined as a result of getting different words correct. Will qualitative differences in the types of response appear between normals and abnormals and between the younger and older normals and abnormals when their scores are based on getting the same words correct? To find the answer, investigation was made of the first 10 words of the Stanford-Binet Vocabulary Test because, practically all the subjects, at all age ranges, answered these words correctly. Analysis of the data was carried out in the same way as was done for the entire list of 45 words, by means of the critical ratio technique.

E. Differences in Qualitative Responses between Normals and Abnormals on First 10 Words of Stanford-Binet Vocabulary Test

Table 12 presents the differences in the qualitative responses of the 135 normals and 135 abnormals of Group I, the 50 normals and 50 abnormals of Group II, and Groups I and II combined to the first ten words of the Vocabulary Test.

Table 12 demonstrates that, at all age ranges, the normals of Group I, Group II (except for the 50-79 age classification), and Groups I and II combined very definitely choose more often the synonym types of definition than do the abnormals. In similar vein the abnormals of Group I, Group II, and Groups I and II combined select significantly more often the use and description types of response, and the demonstration, repetition, illustration and inferior explanation types of answer, at every age classification, than do the normals. In addition, the abnormals, by and large, pick more often the explanation type of response than do the normals.

TABLE 12

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS AND ABNORMALS OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES BY AGE CLASSIFICATIONS (FIRST TEN WORDS OF TEST)

	Group I $N = 38$				Group II $N = 30$				Groups I and II $N = 68$			
Age 15-29	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_n	1.79	.78	1.00	.53	1.41	.86	.91	.53	1.65	.82	.96	.53
SD_a	2.01	1.25	1.12	.88	1.49	1.10	.79	.69	1.81	1.20	1.00	.81
$Mean_n$	7.97	.53	1.00	.24	7.87	.73	.90	.30	7.92	.62	.96	.26
$Mean_a$	6.11	1.45	1.50	.82	6.10	1.73	1.37	.70	6.10	1.57	1.44	.76
$\bar{D}$	1.86	— .92	— .50	— .58	1.77	—1.00	— .47	— .40	1.82	— .95	— .48	— .50
$SE_{\bar{d}}$	.38	.18	.25	.16	.37	.24	.24	.12	.28	.15	.17	.11
$\bar{D}/SE_{\bar{d}}$	4.90**	—5.11**	—2.00*	—3.65**	4.78**	—4.17**	—1.96*	—3.33**	6.50**	—6.33**	—2.82**	—4.55**

	Group I $N = 56$				Group II $N = 16$				Groups I and II $N = 72$			
Age 30-49	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_n	2.17	.89	1.01	.63	.89	.69	1.10	.46	1.95	.85	1.05	.60
SD_a	2.25	1.25	1.05	1.08	2.64	1.22	1.25	1.05	2.35	1.25	1.14	1.08
$Mean_n$	7.59	.68	1.05	.32	7.88	.63	1.13	.31	7.66	.68	1.07	.32
$Mean_a$	5.63	1.79	1.34	1.04	5.06	1.51	2.06	1.13	5.50	1.71	1.50	1.06
$\bar{D}$	1.96	—1.11	— .29	— .72	2.82	— .88	— .93	— .82	2.16	—1.03	— .45	— .74
$SE_{\bar{d}}$	.35	.19	.18	.16	.65	.29	.37	.23	.29	.16	.16	.13
$\bar{D}/SE_{\bar{d}}$	5.94**	—5.84**	—1.61	—4.50**	4.34**	—3.05**	—2.51*	—3.57**	7.45**	—6.44**	—2.69**	—5.69**

—Minus values indicate that the difference is in favor of abnormals.
*Significant at .05 level.
**Significant at .01 level.

TABLE 12 (*continued*)

Age 50-79	Group I $N = 41$ Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Group II $N = 4$ Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Groups I and II $N = 45$ Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_n	2.02	1.01	1.18	.53	1.64	.71	.83	.00	2.12	.98	1.20	.51
SD_c	2.27	1.40	.98	1.22	2.12	1.41	.44	.50	2.30	1.44	.97	1.19
Mean_n	7.39	1.00	1.10	.24	6.25	1.00	2.25	.00	7.22	1.00	1.19	.22
Mean_c	4.59	2.05	1.61	1.59	4.00	3.00	1.25	1.50	4.49	2.11	1.56	1.56
$\bar{D}$	2.80	—1.05	— .51	—1.35	2.25	—2.00	1.00	—1.50	2.73	—1.11	— .37	—1.34
$SE_{\bar{d}}$	.35	.22	.24	.19	1.38	.91	.58	.29	.33	.22	.23	.18
$\bar{D}/SE_{\bar{d}}$	8.00**	—4.77**	—2.13*	—7.11**	1.64	—2.20*	1.72	—5.17**	8.27**	—5.05**	—1.61	—7.44**

Age 15-79	Group I $N = 135$ Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Group II $N = 50$ Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Groups I and II $N = 185$ Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_n	2.02	.93	1.07	.58	1.38	.80	1.05	.49	1.88	.89	1.06	.55
SD_c	2.23	1.38	1.05	1.12	2.09	1.22	1.00	.85	2.23	1.29	1.04	1.07
Mean_n	7.64	.74	1.05	.27	7.74	.72	1.08	.28	7.66	.75	1.06	.23
Mean_c	5.44	1.76	1.47	1.14	5.60	1.76	1.58	.90	5.48	1.77	1.50	1.07
$\bar{D}$	2.20	—1.02	— .42	— .87	2.14	—1.04	— .50	— .62	2.18	—1.04	— .44	— .79
$SE_{\bar{d}}$	.21	.12	.12	.10	.32	.18	.20	.11	.18	.10	.11	.08
$\bar{D}/SE_{\bar{d}}$	10.48**	—8.50**	—5.50**	—8.70**	6.69**	—5.78**	—2.50*	—5.64**	12.11**	—10.40**	—4.00**	—9.88**

—Minus values indicate that the difference is in favor of abnormals.
*Significant at .05 level.
**Significant at .01 level.

F. Differences in Qualitative Responses between Young and Old Normals on First 10 Words of the Stanford-Binet Vocabulary Test

To discover whether significant differences exist in the quality of the definitions given to the first 10 words of the Vocabulary Test by the old as against the young normals, the differences in the mean category scores for the various age levels 15-29, 30-49, and 50-79 were determined, and an evaluation made of the significance of these differences in terms of the standard error of their differences.

The results for the normals of Group I, Group II, and Groups I and II combined appear in Tables 13, 14, and 15.

The findings in Tables 13, 14, and 15 show that for Group I, on the whole, no real differences appear between the types of definition selected by the older as against the younger normals. One exception, however, seems to be in the use and description category which the normals, aged 50-79, choose more often than do the normals, aged 15-29. The results for Group II indicate that the older normals pick more often the explanation type of response than do the younger normals. They also indicate, in contradistinction to Group I, that the younger normals choose more often the demonstration, repetition, inferior explanation, and illustration types of definition than the older normals do. This seeming difference may be explained by the small number of cases in the 50-79 age classification. The data for Groups I and II combined show no significant differences in the qualitative definitions given by the older as against the younger normals. One exception to this finding is that the use and description category, as in Group I, is chosen significantly more often by the 50-79 age classification than by the 15-29 age classification.

G. Differences in Qualitative Responses between Young and Old Abnormals on the First 10 Words of the Stanford-Binet Vocabulary Test

To determine whether any significant differences exist in the quality of definitions given to the first 10 words of the Vocabulary Test between the young and old abnormals, analysis was made of their responses by means of the critical ratio technique.

The results for the abnormals of Group I, Group II and Groups I and II combined are presented in Tables 16, 17, and 18.

TABLE 13

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS (15-29) AND NORMALS (50-49) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES (FIRST TEN WORDS OF TEST)

	Group I *a* 15-29; $N = 38$ *b* 50-49; $N = 56$				Group II *a* 15-29; $N = 30$ *b* 30-49; $N = 16$				Groups I and II *a* 15-29; $N = 68$ *b* 30-49; $N = 72$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	1.79	.78	1.00	.53	1.41	.86	.91	.55	1.65	.82	.96	.55
SD_b	2.17	.89	1.01	.63	.89	.69	1.10	.46	1.95	.85	1.03	.60
Mean_a	7.97	.55	1.00	.24	7.87	.75	.90	.50	7.92	.62	.96	.26
Mean_b	7.59	.68	1.05	.32	7.88	.63	1.13	.31	7.66	.68	1.07	.32
$\overline{D}_{a-b}$	.38	— .15	— .05	— .08	— .01	.10	— .25	— .01	.26	— .06	— .11	— .06
$\widetilde{\sigma}_{a-b}$	.40	.17	.22	.14	.35	.22	.33	.14	.30	.15	.14	.10
$\frac{\overline{D}_{a-b}}{\widetilde{\sigma}_{a-b}}$	.95	— .88	— .23	— .57	— .03	.45	— .70	— .07	.87	— .46	— .79	— .60

—Minus values indicate that the difference is in favor of older group.

TABLE 14

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS (15-29) AND NORMALS (50-79) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES (FIRST TEN WORDS OF TEST)

	Group I *a* 15-29; $N = 58$ *c* 50-79; $N = 41$				Group II *a* 15-29; $N = 50$ *c* 50-79; $N = 4$				Groups I and II *a* 15-29; $N = 68$ *c* 50-79; $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	1.79	.78	1.00	.53	1.41	.86	.91	.53	1.65	.82	.96	.53
SD_c	2.02	1.01	1.18	.53	1.64	.71	.83	.00	2.12	.98	1.20	.51
$Mean_a$	7.97	.53	1.00	.24	7.87	.73	.90	.30	7.92	.62	.96	.26
$Mean_c$	7.39	1.00	1.10	.24	6.25	1.00	2.25	.00	7.22	1.00	1.19	.22
$\overline{D}_{a-c}$	.58	— .47	— .10	.00	1.62	— .27	—1.35	.30	.70	— .38	— .23	.04
$\tilde{\sigma}_{a-c}$	.42	.20	.24	.14	.98	.44	.51	.10	.37	.16	.20	.10
$\frac{\overline{D}_{a-c}}{\tilde{\sigma}_{a-c}}$	1.38	—2.35*	— .42	.00	1.65	— .61	—2.65**	3.00**	1.89	—2.38*	—1.15	.40

—Minus values indicate that the difference is in favor of older group.

*Significant at .05 level.

**Significant at .01 level.

TABLE 15

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN NORMALS (30-49) AND NORMALS (50-79) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES (FIRST TEN WORDS OF TEST)

	Group I *b* 30-49; $N = 56$ *c* 50-79; $N = 41$				Group II *b* 30-49; $N = 16$ *c* 50-79; $N = 4$				Groups I and II *b* 30-49; $N = 72$ *c* 50-79; $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_b	2.17	.89	1.01	.63	.89	.69	1.10	.46	1.95	.85	1.03	.60
SD_c	2.02	1.01	1.18	.55	1.64	.71	.83	.00	2.12	.98	1.20	.51
Mean_b	7.59	.68	1.05	.32	7.88	.63	1.13	.31	7.66	.68	1.07	.32
Mean_c	7.39	1.00	1.10	.24	6.25	1.00	2.25	.00	7.22	1.00	1.19	.22
$\bar{D}_{b-c}$	.20	— .32	— .05	.08	1.63	— .57	—1.12	.31	.44	— .32	— .12	.10
$\sim$ $\sigma_{\bar{D}_{b-c}}$	.42	.17	.22	.14	.97	.45	.56	.10	.39	.17	.20	.11
$\frac{\bar{D}_{b-c}}{\sim}$ $\sigma_{\bar{D}_{b-c}}$	.48	—1.88	— .23	.57	1.68	— .82	—2.00*	3.10**	1.13	—1.88	— .60	.91

—Minus values indicate that the difference is in favor of older group.

*Significant at .05 level.

**Significant at .01 level.

TABLE 16

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN ABNORMALS (15-29) AND ABNORMALS (30-49) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES (FIRST TEN WORDS OF TEST)

	Group I *a* 15-29; $N = 38$ *b* 30-49; $N = 56$				Group II *a* 15-29; $N = 50$ *b* 30-49; $N = 16$				Groups I and II *a* 15-29; $N = 68$ *b* 30-49; $N = 72$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	2.01	1.25	1.12	.88	1.49	1.10	.79	.69	1.81	1.20	1.00	.81
SD_b	2.23	1.25	1.05	1.08	2.64	1.22	1.25	1.05	2.35	1.25	1.14	1.08
Mean_a	6.11	1.45	1.50	.82	6.10	1.73	1.37	.70	6.10	1.57	1.44	.76
Mean_b	5.63	1.79	1.34	1.04	5.06	1.51	2.06	1.13	5.50	1.71	1.50	1.06
$\overline{D}_{a-b}$	.48	— .34	.16	— .22	1.04	.22	— .69	— .43	.60	— .14	— .06	— .30
$\tilde{\sigma}_{a-b}$	.45	.26	.22	.20	.73	.37	.36	.30	.36	.20	.17	.16
$\frac{\overline{D}_{a-b}}{\tilde{\sigma}_{a-b}}$	1.07	—1.31	.73	—1.10	1.42	.60	—1.92	—1.43	1.67	— .70	— .35	—1.88

—Minus values indicate that the difference is in favor of older group.

TABLE 17

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN ABNORMALS (15-29) AND ABNORMALS (50-79) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES (FIRST TEN WORDS OF TEST)

	Group I *a* 15-29; $N = 38$ *c* 50-79; $N = 41$				Group II *a* 15-29; $N = 30$ *c* 50-79; $N = 4$				Groups I and II *a* 15-29; $N = 68$ *c* 50-79; $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_a	2.01	1.25	1.12	.88	1.49	1.10	.79	.69	1.81	1.20	1.00	.81
SD_c	2.27	1.40	.98	1.22	2.12	1.41	.44	.50	2.30	1.44	.97	1.19
$Mean_a$	6.11	1.45	1.50	.82	6.10	1.73	1.37	.70	6.10	1.57	1.44	.76
$Mean_c$	4.59	2.05	1.61	1.59	4.00	3.00	1.25	1.50	4.49	2.11	1.56	1.56
$\bar{D}_{a-c}$	1.52	— .60	— .11	— .77	2.10	—1.27	.12	— .80	1.61	— .54	— .12	— .80
$\tilde{\sigma}_{a-c}$	.49	.30	.22	.24	1.25	.84	.28	.52	.41	.26	.17	.19
$\frac{\bar{D}_{a-c}}{\tilde{\sigma}_{a-c}}$	3.10**	—2.00**	— .50	—3.21**	1.68	—1.51	.45	—2.50*	3.95**	—2.08*	— .71	—4.21**

—Minus values indicate that the difference is in favor of older group.

*Significant at .05 level.

**Significant at .01 level.

TABLE 18

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF THE DIFFERENCES BETWEEN ABNORMALS (30-49) AND ABNORMALS (50-79) OF GROUP I, GROUP II, AND GROUPS I AND II COMBINED ON QUALITATIVE CATEGORIES (FIRST TEN WORDS OF TEST)

	Group I *b* 30-49; $N = 56$ *c* 50-79; $N = 41$				Group II *b* 30-49; $N = 16$ *c* 50-79; $N = 4$				Groups I and II *b* 30-49; $N = 72$ *c* 50-79; $N = 45$			
	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.	Synonym	Use and Description	Explanation	Demonstration Repetition, etc.
SD_b	2.23	1.25	1.05	1.08	2.64	1.22	1.25	1.05	2.35	1.25	1.14	1.08
SD_c	2.27	1.40	.98	1.22	2.12	1.41	.41	.50	2.30	1.44	.97	1.19
Mean_b	5.63	1.79	1.34	1.04	5.06	1.51	2.06	1.13	5.50	1.71	1.50	1.06
Mean_c	4.59	2.05	1.61	1.59	4.00	3.00	1.25	1.50	4.49	2.11	1.56	1.56
$\overline{D}_{b-c}$	1.04	— .26	— .27	— .55	1.06	—1.49	.81	— .37	1.01	— .40	— .06	— .50
$\sigma_{\overline{D}_{b-c}}$	.47	.28	.20	.24	1.40	.88	.41	.39	.45	.26	.20	.22
$\frac{\overline{D}_{b-c}}{\sigma_{\overline{D}_{b-c}}}$	2.21*	— .93	—1.35	—2.29*	.76	—1.69	1.98*	— .95	2.24*	—1.54	— .30	—2.27*

—Minus values indicate that the difference is in favor of older group.
*Significant at .05 level.
**Significant at .01 level.

The findings in Tables 16, 17, and 18 show that for Group I the younger abnormals pick more often the synonym types of definition than do the older abnormals. The older abnormals tend to select more often the use and description category than the younger abnormals do. They definitely choose more often the demonstration, repetition, illustration, and inferior explanation types of response than do the younger abnormals. The data for Group II indicate that the 30-49 age classification tends to select more often the explanation type of answer than either the 15-29 or 50-79 age group, and that the 50-79 age classification chooses more often the inferior explanation, illustration, repetition, and demonstration class of answers than does the 15-29 age group. The results for the abnormals of Groups I and II combined demonstrate that the younger abnormals select significantly more often the synonym types of definition than do the older abnormals. The older abnormals tend to pick more often the use and description kinds of response than do the younger abnormals. They definitely select more often the inferior explanation, demonstration, illustration, and repetition types of response than do the younger abnormals.

A comparison of the qualitative preferences between the normal and abnormal subjects on the first 10 words shows a very parallel picture to the one found for the entire 45 words of the Vocabulary Test. The results indicate the following: (*a*) The normals significantly more often choose the synonym types of response, whereas the abnormals significantly more often select the use and description, explanation, and repetition, demonstration, inferior explanation, and illustration types of definition. (*b*) The younger abnormals pick more often the synonym types of response in contrast to the older abnormals, than do the younger normals when compared with the older normals. (*c*) Both older normals and abnormals tend to select more often the use and description category than do the younger normals and abnormals. (*d*) No clear differences appear between the younger normals and abnormals and the older normals and abnormals in their selection of the explanation type of response. (*e*) The older abnormals significantly more often choose the inferior explanation, demonstration, repetition, and illustration category than do the younger abnormals in contrast to the older normals who do not select this category of response any more frequently than do the younger normals.

H. Findings

Analysis of the results based on the entire 45 words of the Vocabulary Test and on the first 10 words of the Test indicates that they both present a very similar picture. A listing of the findings is summarized below:

1. The normals significantly more often choose the synonym types of definition than do the abnormals at all age classifications.

2. The abnormals significantly more often select the use and description types of definition than do the normals at all age classifications.

3. The abnormals pick the explanation type of definition significantly more often than do the normals at practically all age classifications.

4. The abnormals choose the inferior explanation, demonstration, illustration, and repetition types of definition significantly more often than do the normals at all age classifications.

5. No clear differences are indicated between the younger and older normals in their choice of the synonym types of definition.

6. There is a general trend for the older normals to select more often the use and description types of response than do the younger normals.

7. No distinct differences appear between the younger and older normals in their selection of the explanation type of response.

8. No well established differences appear between the younger and older normals in choosing the demonstration, repetition, inferior explanation, and illustration types of definition.

9. The younger abnormals pick more often the synonym types of response than do the older abnormals.

10. The older abnormals tend to select more often the use and description category of response than do the younger abnormals.

11. No clear differences are evidenced between the younger and older abnormals in their choice of the explanation type of response.

12. The older abnormals significantly more often choose the inferior explanation, demonstration, illustration, and repetition types of definition than do the younger abnormals.

To illustrate the clear differences found between the qualitative responses given to the words of the Vocabulary Test by the normals and the abnormals used in this study, some typical definitions are listed below to the first three words of the Test, i.e., "orange," "envelope," and "straw." To the word "orange" the normals responded with answers of the following kind: "a citrus fruit," "a color," "a fruit that we eat," "a fruit that grows on trees," "a tropical fruit that looks like a ball," etc. The abnormals responded to the same word with definitions of the following kind: "you eat it," "it can be peeled," "what you drink," "it's round like a ball and you eat it," "it's yellow," etc. To the word "envelope" the normals gave replies of the following variety: "a container," "a receptacle for paper," "something to put a letter in," "enclosure for letters for mailing," etc. The abnormals gave defi-

nitions of the following kind to the same word: "a piece of paper you fold," "you write letters," "it's sticky on top so you can paste it down," "to mail," "put mail into," etc. To the word "straw" the normals responded with definitions of the following type: "hay," "dried grass," "dry grain that cattle eat," "food that a cow likes," "from oats after you beat them," etc. The abnormals responded to the same word with definitions of the following type: "grows in the field," "what horses sleep in and it's in a barn," "what you make a hat out of," "you suck it," "straw-hat," etc.

IV. SUMMARY AND CONCLUSIONS

The basic purpose of this study was to investigate the hypothesis that the reason the vocabulary scores of mentally impaired persons show a minimal amount of change from their pre-impaired level, and the reason the vocabulary of older persons does not decline as do other tests is because the present scoring system used assumes all right answers are of equal value and does not take into account the quality of the definition response.

A. Summary

The group used for study consisted of 370 normal and abnormal subjects, male and female, ranging from 15 to 80 years of age, and representing an above average group in terms of mental ability. The cases were divided into a main group of 135 normals and 135 abnormals matched on the variables of age, education, and score correct on the Stanford-Binet Vocabulary Test, and another one of 50 normals and 50 abnormals matched on the factors of age and score correct on the Stanford-Binet Vocabulary Test. The abnormal subjects were secured at St. Elizabeths Hospital in Washington, D. C. Both normal and abnormal subjects came essentially from the Eastern seaboard states. Both groups were also divided into three main age divisions: 15-29, 30-49, and 50-79. All their verbatim responses, i.e., each definition to every word of the 45 words of the Test, were qualitatively analyzed by means of a five-fold qualitative category system. The reliability of these categories was found to be very high for each one of the 45 words of the Test. Statistical analysis, utilizing the critical ratio technique, was made of the qualitative differences in response between the normal and abnormal subjects, and between the younger and older normals, and younger and older abnormals.

Even though the normal and abnormal subjects were matched on total score correct, it was possible for similar scores to result from getting different words correct. To make certain that differences in qualitative modes of approach were being secured rather than a possible differential intelligence factor, a similar analysis as was done with all 45 words of Test was carried out for the first 10 words of the Test because these were answered correctly by practically all the subjects. Examination of the results demonstrated very identical outcomes for both the 45 words and 10 words. The principal findings were as follows:

1. The normal subjects, at all age classifications, selected the synonym types of response significantly more often than did the abnormals.

2. The abnormal subjects, at all age classifications, significantly more often chose the use and description types of response than did the normals.

3. The abnormal subjects, at all age classifications, picked the explanation type of response significantly more often than did the normals.

4. The abnormals, at all age classifications, significantly more often selected the inferior explanation, demonstration, illustration, and repetition types of response than did the normals.

5. No clear differences were indicated between the younger and older normals in their choice of the synonym types of definition.

6. The older normals tended to pick more frequently the use and description types of response than did the younger normals.

7. No well established differences appeared between the younger and older normals in their selection of the explanation type of response.

8. No significant differences appeared between the younger and older normals in picking the demonstration, repetition, inferior explanation, and illustration types of definition.

9. The younger abnormals significantly more often selected the synonym types of response than did the older abnormals.

10. The older abnormals tended to choose more often the use and description types of response than did the younger abnormals.

11. No distinct differences were evidenced between the younger and older abnormals in their choice of the explanation type of response.

12. The older abnormals significantly more often selected the inferior explanation, illustration, repetition, and demonstration types of definition than did the younger abnormals.

B. Conclusions and Implications

Significant differences have been demonstrated in the qualitative types of response given by normals when compared with abnormals similar in background. Analysis of the verbatim responses shows that the abnormals significantly more often select the use and description types of response, and the demonstration, illustration, repetition, and inferior explanation kinds of definition. Binet and Simon (4) and Terman (42) both found that the use type of definition response was most frequently employed by children eight years of age and younger. Green (21) discovered that the use, description, repetition, and demonstration types of definition response were most characteristic of young children. Feifel and Lorge (13), employing the same five-fold qualitative system used in this study, demonstrated that children between the ages of six and nine significantly more often chose the

use and description types of definition, as well as the inferior explanation, demonstration, repetition, and illustration types of response when compared with older children, aged 10 to 14, who in their turn, significantly more often chose the synonym types of response. Obviously, the qualitative mode of approach used by the abnormals in this study is very similar to that of the younger child. This appears to support the outlook of research workers like Vigotsky (46) who contended that in schizophrenia a regression takes place in the type of thought used to one which is fundamentally similar in most essentials to that of children, and Goldstein (18) who indicated that both schizophrenics and organics act on a much earlier level than normal persons with the same intelligence and social status.

Piaget (30) stressed the intimate connection between words and concept formation. Along this line of evidence, the present study supports the work of Bolles and Goldstein (6) who found that abnormal persons tended to respond in a concrete manner to most situations rather than abstractly, and the findings of Kasanin (25) who stated that in abnormals there is a reduction in conceptual thinking and in their capacity to think abstractly. Analysis of the verbatim qualitative definitions given by our abnormals shows that they tend to perceive words as concrete ideas and do not generalize from the particular. Their power to abstract relations is also very weak. To the abnormal, "scorch" means "hot" rather than "to burn," "straw" is "yellow" rather than "dry grass," etc. This concrete manner of approach utilized by the abnormal again points up his close relation to the child since the child in his definition of words emphasizes the particular or isolated aspect rather than the categorical or "class" feature of the meaning. The abnormal person's manner of approach, then, as evidenced through his definitions of words strongly evidences concrete outlook, a matter-of-fact approach in which things have personal rather than symbolic value, and paucity of ability to grasp the idea of classification and abstraction. It should be remembered, however, that no particular type of definition response is the exclusive property of any group, or, for that matter, any specific group.

The findings challenge one of the main assumptions underlying the Babcock test for mental deterioration, namely, that the reason vocabulary fails to show decline in contrast to the reduction in score on other types of tests is that vocabulary is an old habit compared to other abilities involving more recent associations. Rather they sustain the Yacorzynski hypothesis that the reason the vocabulary scores of abnormal subjects show a minimal amount of change from presumptive pre-impaired level is that the subject can solve the problem, or give the correct response, on the basis of an "easier definition"

even when the mechanisms called for in more difficult conceptual organization are no longer available. The relatively lower standards for acceptable achievement on the vocabulary test are well suited to the psychotic's generally lowered efficiency. Since the psychotic is less penalized by approximate answers an apparent but contaminated superiority in performance is indicated in the abnormal's vocabulary scores when compared with other test scores.

Vocabulary functioning has long been recognized as associated with the development of intelligence and general mental functioning. The data presented in this study indicate that it is also closely related to the impairment of mental functioning when vocabulary responses are qualitatively analyzed. Because the manner of approach and the method of procedure used by the abnormal person in defining words is as important in understanding his mental processes as his successes and failures on the test, some measure of the loss of mental functioning by the subject can be secured by studying the patterns of response given by him. Since the first 10 words of the Revised Stanford-Binet Vocabulary Test are correctly answered by practically all subjects, it may serve as a possible clinical tool in suggesting the degree of mental impairment present.

The results, in general, do not establish the clear-cut differences in qualitative types of response for the older adults when compared with the younger adults, as was discovered between the normal and abnormal subjects. This seems to corroborate the findings of Fox (14). However, there are suggestions that the types of response employed by the older adults tend to be similar to those used by children. This tendency is much more pronounced and statistically verifiable when older abnormals are compared with younger abnormals than when older normals are compared with younger normals. Seemingly, the relation between age and regression to earlier levels of qualitative types of definition is stronger in abnormals than in normals.

REFERENCES

1. Ackerman, S. B. Vocabulary and mental deterioration in senile dementia. *J. Abn. & Soc. Psychol.*, 1944, **39**, 393-406.
2. Arrington, R. E. Interrelations in the behavior of young children. *Child Devel. Monog.*, 1932, No. 8.
3. Babcock, H. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, 1930, **18**, No. 117.
4. Binet, A., & Simon, T. The Development of Intelligence in Children. Publications of the Training School at Vineland, New Jersey, 1916. Pp. 336.
5. Bolles, M. M. The basis of pertinence. *Arch. of Psychol.*, 1937, **30**, No. 212.
6. Bolles, M., & Goldstein, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42-65.
7. Capps, H. M. Vocabulary changes in mental deterioration. *Arch. of Psychol.*, 1939, **34**, No. 242.
8. Chambers, W. G. How words get meaning. *Ped. Sem.*, 1904, **11**, 30-50.
9. Christian, A. M., & Paterson, D. G. Growth of vocabulary in later "maturity." *J. of Psychol.*, 1936, **1**, 167-169.
10. Davidson, M. A study of schizophrenic performance on the Stanford-Binet scale. *Brit. J. Med. Psychol.*, 1937, **17**, 93-97.
11. ————. Studies in the application of mental tests to psychotic patients. *Brit. J. Med. Psychol.*, 1939, **18**, 44-52.
12. Dolch, E. W. Reading and Word Meanings. Boston: Ginn, 1927. Pp. 129.
13. Feifel, H., & Lorge, I. Qualitative differences in the vocabulary responses of children. (To be published.)
14. Fox, C. Vocabulary ability in later maturity. *J. Educ. Psychol.*, 1947, **38**, 482-492.
15. Garrett, H. E. Statistics in Psychology and Education. (3rd Ed.) New York: Longmans, 1947. Pp. 493.
16. Gilbert, J. G. Mental efficiency in senescence. *Arch. of Psychol.*, 1935, **27**, 188.
17. Goldstein, K. The Organism. New York: American Book, 1939. Pp. 533.
18. ————. Methodological approach to the study of schizophrenic thought disorder. *In* J. S. Kasanin (*Ed.*), *Language and Thought in Schizophrenia*. Berkeley and Los Angeles: Univ. California Press, 1946. (Pp. 17-40.)
19. Gottschalk, A. J. The institutional psychotic epileptic. *Amer. J. Psychiat.*, 1942, **98**, 839-841.
20. Gray, W. S., & Holmes, E. The Development of Meaning Vocabularies in Reading. Publications of the Laboratory Schools of The Univ. of Chicago, No. 6, 1938. Pp. 140.
21. Green, H. J. A qualitative method for scoring the vocabulary test of the new revision of the Stanford-Binet. Unpublished Master's thesis, Stanford Univ., 1931.
22. Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1936. Pp. 566.
23. Jastak, J. Psychometric patterns of state hospital patients. *Delaware St. Med. J.*, 1937, **9**, 7-10.
24. Jones, H. E., & Conrad, H. S. The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monog.*, 1933, **13**, 223-298.
25. Kasanin, J. S. The disturbance of conceptual thinking in schizophrenia. *In* J. S. Kasanin (*Ed.*), *Language and Thoughts in Schizophrenia*. Berkeley and Los Angeles: Univ. California Press, 1946. (Pp. 41-49.)
26. Kendig, I., & Richmond, W. V. Psychological Studies in Dementia Praecox. Ann Arbor, Michigan: Edwards Brothers, 1940. Pp. 211.

27. KIRKPATRICK, E. A. A vocabulary test. Quoted in W. S. Gray and Eleanor Holmes, *The Development of Meaning Vocabularies in Reading*. Chicago: Univ. Chicago Press, 1938. P. 17.

28. MALAMUD, W., & PALMER, E. M. Intellectual deterioration in the psychosis. *Arch. Neurol. & Psychiat.*, 1938, **39**, 68-81.

29. MARX, B. A study of the first fifty words of the Stanford-Binet Vocabulary. Unpublished Master's thesis, Stanford Univ., 1928.

30. PIAGET, J. The Language and Thought of the Child. New York: Harcourt, 1926. Pp. 246.

31. ———. Judgment and Reasoning in the Child. New York: Harcourt, 1928. Pp. 260.

32. RABIN, A. I. The relationship between vocabulary levels and levels of general intelligence in psychotic and non-psychotic individuals of a wide age-range. *J. Educ. Psychol.*, 1944, **35**, 411-422.

33. ———. Vocabulary and efficiency levels as functions of age in the Babcock method. *J. Consult. Psychol.*, 1947, **11**, 207-211.

34. RAPAPORT, D. Manual of Diagnostic Psychological Testing. New York: Josiah Macy, Jr. Foundation, 1944. (Vol. II.) Pp. 239.

35. ROE, A., & SHAKOW, D. Intelligence in mental disorder. *Ann. N. Y. Acad. Sci.*, 1942, **42**, 361-490.

36. SCHWARTZ, R. Mental deterioration in dementia praecox. *Amer. J. Psychiat.*, 1932, **12**, 555-560.

37. SHAKOW, D., & GOLDMAN, R. The effect of age on the Stanford-Binet vocabulary score of adults. *J. Educ. Psychol.*, 1938, **29**, 241-256.

38. SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. of Psychol.*, 1940, **9**, 371-377.

39. SORENSON, H. Mental ability over a wide range of adult ages. *J. Appl. Psychol.*, 1933, **17**, 729-741.

40. SWARD, K. Age and mental ability in superior men. *Amer. J. Psychol.*, 1945, **58**, 443-479.

41. TERMAN, L. M., & CHILDS, H. G. A tentative revision and extension of the Binet-Simon measuring scale of intelligence. *J. Educ. Psychol.*, 1912, **3**, 198-208.

42. TERMAN, L. M. The Measurement of Intelligence. Boston: Houghton Mifflin, 1916. Pp. 362.

43. ———. The vocabulary test as a measure of intelligence. *J. Educ. Psychol.*, 1918, **9**, 452-466.

44. TERMAN, L. M., & MERRILL, M. A. Measuring Intelligence. Boston: Houghton Mifflin, 1937. Pp. 461.

45. THORNDIKE, R. L., & GALLUP, G. H. Verbal intelligence of the American adult. *J. Gen. Psychol.*, 1944, **30**, 75-85.

46. VYGOTSKY, L. S. Thought in schizophrenia. *Arch. Neurol. & Psychiat.*, 1934, **31**, 1063-1077.

47. WATTS, A. F. The Language and Mental Development of Children. London and Toronto: Harrap, 1944. Pp. 354.

48. WECHSLER, D. The Measurement of Adult Intelligence. (3rd Ed.) Baltimore: Williams & Wilkins, 1944. Pp. 258.

49. WEISENBURG, T. H., ROE, A., & MCBRIDE, K. E. Adult Intelligence. New York: Commonwealth Fund, 1936. Pp. 155.

50. WILLOUGHBY, R. R. Family similarities in mental test abilities. *Genet. Psychol. Monogr.*, 1927, **2**, 235-277.

51. WITTMAN, P. The Babcock deterioration test in state hospital practice. *J. Abn. & Soc. Psychol.*, 1933, **28**, 70-84.

52. YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock deterioration test. *Psych. Bull.*, 1940, **37**, 425-426.

53. ———. An evaluation of the postulates underlying the Babcock deterioration test. *Psychol. Rev.*, 1941, **48**, 261-267.

Published as a separate and in *Genetic Psychology Monographs*, 1949, **39**, 205-278.

THE RELATIVE EFFECTIVENESS OF MOTION AND STILL PICTURES AS STIMULI FOR ELICITING FANTASY STORIES ABOUT ADOLESCENT-PARENT RELATIONSHIPS*[1]

Division of Social Sciences, University of Chicago

PAUL E. EISERER

Acknowledgments 207

I. Introduction 209

II. Procedure 215

III. Story productivity and associated relationships 227

IV. Methodological aspects of the study 233

V. Content characteristics of the stories and pictures 261

VI. Reliability of the study 269

VII. Conclusions, implications for personality study, and suggestions for further research 273

References 277

*Received in the Editorial Office on January 13, 1949.

[1]This publication represents the substance of a Ph.D. dissertation accepted by the Division of the Social Sciences, University of Chicago, June, 1948.

ACKNOWLEDGMENTS

This investigation has been accomplished through the coöperation of many people. The writer expresses his gratitude to Professors Stephen M. Corey, Robert J. Havighurst, William E. Henry and Ralph W. Tyler for their continued interest and creative support. Dr. Cyril Hoyt of the University of Minnesota gave invaluable guidance in experimental design and statistical analysis. For their assistance as judges and producers of the motion pictures I am indebted to Dr. Frank Costin of the University of Illinois, Dr. Frank Estvan of the University of Southern California, and Messrs. Newton Calhoun, Eugene Dutton, and Alexander Howard of the University of Chicago.

PAUL E. EISERER

University of Oregon
Eugene, Oregon

I. INTRODUCTION

A. Background of the Problem

The purpose of this study is to explore experimentally some of the uses of the motion picture as a projective method for studying personality. Projective methods were first defined by Frank (15) as follows:

> A projective method of the study of personality involves the presentation of a stimulus-situation designed or chosen because it will mean to the subject, not what the experimenter has arbitrarily decided it should mean (as in most psychological experiments using standardized stimuli in order to be "objective") but rather whatever it must mean to the personality who gives it, or imposes upon it, his private, idiosyncratic meaning and organization.

The motion picture has heretofore been used chiefly for three purposes. (*a*) The most common use has been as a recreational medium. (*b*) The motion picture has been developed in recent times as an instructional aid in achieving educational objectives. (*c*) The motion picture has been employed to record behavior for psychological study and analysis. This is exemplified by the work of Gesell in child study.

Considerably less work has been done using motion pictures as specific test items or stimuli under standardized conditions. A significant attempt in this direction was made by Dysinger and Ruckmick (12) as part of the Payne Fund Studies on Motion Pictures and Youth in 1933. They used a physiological measure to study the emotional responses of children to a variety of motion picture situations. But it was not until World War II that the first comprehensive and systematic utilization of the motion picture as a testing technique was made. This research by the Army Air Forces (28) was aimed at the development of aptitude and proficiency tests.

The motion picture has never been used as a test in the field of personality study. This research, then, may be viewed as an exploratory effort in the general field of personality study as well as in the more specialized area of projective techniques. The point of departure for investigating the potentialities of motion pictures as projective instruments was suggested by the current extensive use of a personality test in which still pictures constitute the stimuli. This is the Thematic Apperception Test, developed and described by Morgan and Murray (27). An essential characteristic of these pictures is that they depict ambiguous situations about which a person can relate stories. The presumption is that in so doing the teller reveals more about himself than about the pictures.

The basic technique employed in this research is to compare motion pictures as projective stimuli with equivalent still pictures in eliciting stories from which inferences could be made. In order to do this, two important technical problems had to be solved. (*a*) Motion pictures which satisfy accepted criteria for use as projective instruments had to be found or developed. (*b*) A procedure for developing still picture equivalents of the motion pictures had to be worked out. How this was done is described in Section II.

In addition to these methodological questions, the investigator's interest in adolescence led to the idea that the study might be focussed on the particular area of parent-adolescent family relationships. This had a bearing on the kinds of picture situations used and is discussed in Section II.

The investigator was originally led to consider the possibility that motion pictures might be more productive than still pictures in stimulating significant psychological material for two reasons. (*a*) It seemed reasonable to suppose that the motion picture, because of its movement properties, can be characterized as more life-like, more realistic, and more inciting than still pictures. In everyday experience, motion pictures seem closer to life than still pictures. Our empirically derived impression is supported by a discussion of the psychology of motion pictures in which Dashiell (7) shows how motion pictures approach the reality of action.

> The enormous success of moving pictures is based on their power to arouse an illusory perceiving of action. The screen, although actually in darkness many times per second, is able to stimulate the audience much as do the continuous movements of actors in person-arousing and holding attentive postures, touching off emotional tendencies, exciting thinking behavior. The situation presented is discontinuous and the observers are not actually looking at objects in motion, but the successive pictures obtained by instantaneous photography follow each other so rapidly, while their actual movement in this succession before the projector is being concealed by a shutter, that the same responses are elicited as though the scenes of things and people were actually moving.

(*b*) Perception of motion is a qualitatively different experience than the perception of a still picture. Everyone is familiar with the fact that if a series of still pictures is presented in rapid succession the observer will perceive motion. The first device for demonstrating this phenomena is credited to Plateau (43), a Belgian physicist, about the middle of the nineteenth century. Psychologists, however, until 1912 were unsuccessful in dealing with the problem of visually perceived movement. Boring (5) says, "pri-

marily the difficulty lay in the fact that psychologists thought of perceived movement as involving a series of sensations, a view that arises out of sensationistic elementarism." In 1912 Wertheimer (22), generally recognized as one of the founders of the Gestalt school of psychology, carried on experimentation which led to the discard of a purely summative theory of sensory experience. The illusion of movement which is created by successive exposures of non-moving stimuli under specified conditions of spatial separation, temporal interval, and intensity was called the phi-phenomenon. The significance of this phenomenon is well stated by Warren and Carmichael (41).

> This phenomenon has attracted considerable interest because it is important in itself and also because it shows that perception is more than the sum of any such elements as sensations. If the two successive presentations that call forth the experience of movement are analyzed so that they are viewed separately, the whole experience of movement vanishes. One sees only two separate stimuli. This total experience has, therefore, been called a configuration. It is an example of the fact that in perception the whole is qualitatively different from a mere adding together of the parts.

After the present research had been started an interesting publication (28) appeared in which the possibilities of using motion pictures in personality study were discussed. It was pointed out that two different and distinct kinds of movement perception may be induced on a screen, namely, movement of *objects* and movement of the *observer himself*. This distinction is made clearer by the following illustration:

> For example, one has a greater tendency to experience what the paratrooper experiences as he makes his jump if the camera looks out and down through the door of the plane than is the case if the camera takes the point of view of a mere onlooker or observer. If the camera moves and shifts its view appropriately the onlooker can be made to identify himself with an active participant in the situation (28).

Insofar as such a differentiation can be maintained, this investigation is confined to the movement of objects in the motion picture although it is recognized that the possibility of projecting the observer himself into a situation by camera techniques is an intriguing idea. A discussion of the unusual capacity of the motion picture to induce observer participation is concluded with this statement. "The usefulness of this characteristic of motion pictures appears to be promising in the field of personality test construction, the tendency to participate in the motion picture scene is a fact which could usefully be exploited in work with projective methods . . ." (28).

B. Framework of the Study

The chief question for examination in this investigation is, how do the motion and still picture compare as instruments for obtaining material which can provide the basis for inferences about the intra-family relationships of adolescents? The first requirement for answering this question was to find motion pictures which could serve as projective instruments. Because there were no existing motion pictures which could practicably be used for this purpose, pictures had to be developed especially for this experiment. At many points the chief basis for determining the course to take was the investigator's judgment. From many possible motion picture techniques one had to be selected. From the almost infinite number of situations depicting parent-adolescent relationships that could be portrayed selection had to be made.

As a first step six motion pictures were developed and used in a pilot study. After it was found that motion pictures which would reasonably satisfy criteria for projective instruments could be developed, more specific questions were formulated. The following questions, which reflect the methodological aspects of the research, gave direction to the experimental design used:

> Does the motion or still picture elicit more expressions of a subject's feelings?
>
> What are the differential effects of various stimuli situations?
>
> What is the effect of the order of presentation of groups of stimuli?
>
> How shall the specific motion pictures and their still picture equivalents be grouped for presentation?
>
> How shall the subjects be assigned to the different groups of stimuli and orders of presentation?
>
> How shall responses be taken? In what form?

In order to provide answers to these questions the following steps, described in detail in Section II, were taken:

1. A series of motion pictures was developed.
2. Equivalent still pictures were constructed.
3. Motion and still pictures were combined into two batteries according to stated criteria.
4. Orders of presentation of motion and still pictures were reversed in each battery for various respondents.
5. The respondents were assigned to four experimental presentations of stimuli.
6. A wire recorder was used to obtain all verbalized responses.

To answer the questions raised by this study it was necessary to quantify the data in a fashion that was meaningful psychologically. Consequently

methods of analysis were used which had some degree of acceptability in the thematic apperception field of inquiry. The chief statistical method used is to evaluate group differences by analysis of variance techniques after several different quantifications of the data have been made. The quantification procedures and analysis of variance technique used are described in Section III.

C. Scope of the Study

The present study has been delimited in the following ways: (*a*) The range of situations that might be depicted in motion or still pictures is probably infinite. Only a limited number could be developed and used within the scope of this study. Murray (29) used hundreds of pictures in order to find the 30 which comprise the Thematic Apperception Test. The investigator conceived many more situations than were used in this study. Fifteen situations were actually photographed from which 12 were selected for final use. (*b*) The kinds of motion picture techniques that might be used to answer the major questions of this study are more extensive than those tried by the investigator. A silhouette type of picture was finally selected although animated cartoons, regular photography, special effects, and trick photography were other possibilities. (*c*) A direct comparison between the current Thematic Apperception Test of Murray (31) and the motion pictures developed cannot be made. The still pictures used cannot be construed as equivalent to the pictures of the *TAT*. Only one card, No. 4 in Murray's battery, is in silhouette form, whereas all of the pictures in the present study are black and white silhouettes. (*d*) The responses consist of stories which subjects made up in reaction to the stimuli presented. This method is employed widely in clinical practice and research. It represents a selection from the total behavioral reactions to the stimuli. (*e*) The assumption is made that the data collected by the two methods represent expressions of fantasy by the subjects. The confidence with which this assumption can be made *depends upon the extent to which the construction of stimuli meets adequately* the standards of pictures which generally are considered successful in eliciting fantasy stories. It depends also upon meeting the requirements of good testing procedures, namely, good rapport of subject with administrator and comprehension by the subject of what he is expected to do.

D. Plan of Presentation

This section has included the background of the problem, the framework and limits of the research. Section II describes the procedures followed in

developing the motion and still pictures and methods of collecting the data. Sections III, IV, and V set forth the hypotheses to be tested, the methods of analysis used to test them, and the results obtained. Section VI presents evidence about the reliability of the study. Section VII presents conclusions drawn from the study, the implications for personality research, and suggestions for further investigation.

II. PROCEDURES

A. Construction of the Techniques

This section describes the procedures used to develop the motion and still pictures, how the test batteries were constructed, and the operations involved in the collection of data.

1. *General Considerations*

Exploration of the possibilities for constructing motion pictures that would satisfy an important criterion of projective techniques—the ambiguity of the stimulus—led to this question: What kind of photography would be most satisfactory? Initial efforts were made to achieve ambiguity by underexposing and overexposing the picture. It seemed that both of these methods represented unrealistic photography and would probably be uninteresting to subjects.

During the examination of different kinds of photography, the experimenter came upon a shadowgraph process that had been used to produce historical films by a school in Vancouver, British Columbia. The shadowgraph technique is a silhouette process in which figures are shown in black against a white background. The actions depicted seemed quite clear but no facial features could be discerned. More important, the process seemed authentic and realistic, that is, the photography was genuine, without such distortions as under or overexposure. The principles for producing still picture silhouettes were adapted to the production of movies experimentally.

While the shadowgraph process is not the only technique that might be used to test the major hypotheses of this study, the investigator felt it worthy of trial because it seemed to have the following desired characteristics: (*a*) The pictures would be realistic, (*b*) facial features could be made more neutral than in most picture techniques, (*c*) the motions depicted could be made to stand out in clear relief, (*d*) ambiguity and simplicity could be readily accomplished, (*e*) the production would be relatively inexpensive.

The shadowgraph process might have the following disadvantages: (*a*) The lack of familiarity of the subjects with this kind of stimulus might create difficulties. (*b*) the relative absence of facial cues might make it more difficult for subjects to relate themselves to the stimuli.

2. *Development of the Motion Pictures*

Research on pictorial methods for stimulating imaginative processes has indicated that the best results are obtained when certain criteria are applied in

the construction of pictures. As a result of experimental investigation, Symonds (38) concluded that the following criteria are important for the selection of pictures for stimulating fantasy production in adolescents: (*a*) Pictures should be ambiguous, (*b*) pictures should include some character with whom the subject can identify, (*c*) pictures should have a minimum of detail, (*d*) situations depicted should be incomplete, episodic.

In order to satisfy the criterion that the pictures include some character with whom the subjects might more readily make identifications boys were used both as participants in a majority of the pictures and as subjects for testing.

A decision was made by the writer to photograph a large variety of situations from which the most promising pictures could be selected for use in the final investigation. Consequently, a total of 15 motion pictures were developed. Family members, such as father, mother, daughter, and son, were represented in different combinations and in varied kinds of interactions.

Experimentation showed that the following method gave the best photographic results. The situations to be taken were enacted in front of a white screen. The lights were placed behind the screen. Except for the lights behind the screen the room was darkened completely. The camera was placed in front of the actors. A 16 mm. camera was used throughout to take pictures at 16 frames per second. A stop watch was used to time each sequence.

Since the actions to be portrayed were relatively simple, it was found that amateurs could serve as actors in the short dramas. Graduate students were used as parents and high school students as boys and girls. Before shooting a sequence, each actor was given detailed instructions and each scene was rehearsed several times.

3. *Preliminary Tryout of Motion Pictures*

To test the hypothesis that the motion picture sequences constructed were capable of eliciting fantasies, the six pictures were administered to a group of students. Since a large number of reactions were more desired than a thorough clinical examination of one individual, 19 eighth-grade boys were selected on the basis of their availability and proximity in age to the group to be used in the final investigation. Responses were written by the subjects.[2] In most instances, the subjects finished writing about each picture within five minutes.

[2] The writer is aware of the opinion of workers in the field that the writing of responses by subjects may impose a more inhibiting and censoring effect on the stories than if stories were spoken (25, 10). This did not seem to be a serious disadvantage for testing the hypothesis stated above.

These data were analyzed with respect to the following questions: (*a*) Are a variety of themes expressed? (*b*) Do the stories have plot and action rather than description? (*c*) Do the stories represent fantasy as indicated by a departure from the physical aspects of the stimuli? The analysis indicated that these stories met the above criteria sufficiently to warrant more extended investigation of the use of motion pictures as a projective technique.

In order to get a wider range of stimuli from which to select pictures for the final test battery, more motion pictures had to be developed. In the development of additional sequences, new constellations of family members such as man-girl, man-woman-boy-girl, man-girl-boy, boy-girl, and man-woman were introduced and the variety of interactions was extended to get a fairly even balance between pictures that were ostensibly neutral in tone and those with apparent positive and negative emotional tones. As a result of this preliminary investigation, nine additional motion picture sequences were developed, making a total of 15 from which a selection could be made for the experiment. These 15 motion pictures ranged in length from 10 to 21 seconds. The investigator made a judgment about the probable feeling tone of each picture as positive, negative, or neutral. The essential characteristics of the 15 motion pictures are given in Table 1.

TABLE 1
ESSENTIAL CHARACTERISTICS OF FIFTEEN PICTURES DEVELOPED

Number	Persons included	Projection time (in seconds)	Tone*	Brief description
1	Woman-Boy	19	*N*	Boy seated, woman standing
2	Boy	17	*U*	Alone-standing
3	Woman-Man-Boy	14	*N*	Seated at dinner table
4	Woman-Boy	16	*N*	Woman seated, boy standing
5	Woman-Man-Boy	16	*N*	Man seated, woman-boy standing
6	Man-Boy	15	*U*	Man-boy seated
7	Man-Woman	12	*U*	Both seated, man with paper, woman knitting
8	Man-Woman	12	*P*	Woman seated, man standing
9	Man-Girl	15	*P*	Man seated, girl seated before him
10	Woman-Boy	12	*N*	Woman seated knitting, boy standing
11	Boy-Girl	14	*P*	Boy-girl seated
12	Boy-Girl	11	*N*	Girl seated, boy standing
13	Man-Woman-Boy-Girl	21	*P*	All seated except boy
14	Man-Girl-Boy	13	*N*	Man-girl seated, boy standing
15	Woman-Boy	10	*U*	Woman-boy standing

*This description represented the experimenter's judgment about the probable feeling tone of the picture—*P*, positive; *N*, negative; *U*, undetermined or neutral.

4. *Development of the Still Pictures*

The development of a still picture equivalent to each motion picture sequence presented three problems. (*a*) A technical process had to be found for constructing each still picture, so that when projected on a screen, the image would be approximately the same size as the moving picture images. (*b*) A method had to be devised for determining which single frame in a motion picture sequence was most nearly equivalent to the total sequence. (*c*) The criteria for making judgments about the equivalence of stimuli had to be defined.

Several procedures for making a selected single frame into a picture for projection were considered. After considering relative costs, time required, and technical difficulties in obtaining an image as clear and precise as the movie image, the following procedure was adopted. The selected single frame was cut from the movie sequence and mounted on a 2 x 2 slide. This picture was projected through a lens powerful enough to enlarge it to proportions equivalent to the motion picture images.

Since the motion picture sequences were shot at 16 frames a second, each sequence contained more than 200 frames. It became necessary to determine which single frame could best represent the total sequence. The 15 sequences differed in amount and kind of movement, and it became apparent that the simpler the motion picture, the more obviously could one frame represent the totality. In this context simplicity implied a minimum of movement and action. In motion picture No. 13, for example, since the only change in position occurred when the boy entered the scene and began talking while the other three people in the picture kept the same positions throughout the sequence, it was comparatively simple to select a frame because all of them were so similar. However, where persons moved around, made gestures, entered or left the scene, the task became more difficult.

The jury method was adopted for obtaining judgments about which single frame was most typical of each movie sequence. Six judges, *A-B-C-D-E-F*, were used. All of them performed the judging operations before any results were analyzed to determine the extent of agreement and disagreement. All judges were graduate students who were familiar with the purpose of the study. Judges *A* (the investigator), *C*, and *D* had previous experience with thematic apperception tests. Judges *A*, *B*, *D*, and *E* participated directly in the construction of the motion pictures. The operation was performed by each judge individually.

The method used by the judges was to describe the particular scene selected

by them, rather than to select one from a series of scenes described by the investigator. In analyzing the results the investigator set up categories for each sequence, each category representing a different aspect of the total sequence. The picture was projected as many times as necessary to obtain an accurate, detailed description of every scene which differed significantly from others in the sequence. Each sequence was described to include (*a*) the beginning scene; (*b*) all action, movement, and positional changes; (*c*) the ending scenes. It can be seen from Table 2 that the number of categories

TABLE 2
AGREEMENT AMONG JUDGES IN DETERMINING EQUIVALENCY OF STILL AND MOTION PICTURES

Motion picture sequence number	Number of categories	Judgments in same category	Judgments in different category	Percentage of agreement
1	6	*ABCDE*	*F*	83
2	6	*	------	50
3	7	*ABCDEF*	None	100
4	9	*ABDEF*	*C*	83
5	11	*ACDEF*	*B*	83
6	6	*ABCDEF*	None	100
7	5	*ABDEF*	*C*	83
8	5	*ABCEF*	*D*	83
9	5	*ABCDEF*	None	100
10	7	*ABCDEF*	None	100
11	6	*ABCDEF*	None	100
12	8	*ABCEF*	*D*	83
13	4	*ABCDEF*	None	100
14	6	*ABDEF*	*C*	83
15	5	*ABCDEF*	None	100

*Selections of judges *ACD* fell in one category, of judges *BEF* in another category.

ranged from 4 to 11. The scenes described by the judges were then classified into the appropriate categories. In many instances the descriptions of the judges were quite similar to the description of the category. This made classification relatively easy. In some instances scenes at the beginning and at the end of a sequence were about the same, with different scenes between them. When one of these relatively similar scenes was selected, as happened infrequently, it could be classified in the same category, for there was no difference for the purpose of this procedure.

The extent of agreement of the judges was impressive (see Table 2). The six judges were in complete agreement in 7 out of 15 cases. In seven cases five out of six judges agreed. In one case the selection of three judges fell into one category, the selection of the other three judges fell into one category. The basis of disagreement was whether the boy in the picture

should place his hands behind his back or place one hand on his head. Only in this case did Judges *A* and *E* disagree. It is interesting that these judges agreed completely in all the other cases, the only judges to do so, for they were most closely associated in the development of the motion pictures. The single frames thus selected were cut out of the movie sequence and mounted on 2 x 2 slides.

5. *Making the Test Batteries*

As a result of the procedures that have been described, 15 brief motion picture sequences and 15 comparable still pictures were made available. Then an experimental design was developed to accomplish two objectives: (*a*) Meet the requirements of time available (one school period), minimize fatigue factors, and maintain the interest of the subjects. (*b*) Make possible the collection of data that would represent an adequate sampling of the subject's behavior with respect to the major hypotheses of the study. The following plan was adopted:

> Develop a test battery that would include both motion and still pictures. The complete series would be given to each subject. Two test batteries would be constructed, each to include a series of motion and still pictures. The still pictures from Battery *A* would be placed with Battery *B* movies and vice versa.

This plan involved the assumption that Battery *B* still pictures were equivalent to Battery *A* motion pictures and that Battery *A* still pictures were equivalent to Battery *B* motion pictures. This assumption appeared reasonable since the selection and grouping of motion and still picture stimuli were made in accordance with criteria which supported the view that both batteries of motion pictures contained many common elements.

In placing the various pictures in the two batteries, the following criteria were considered: (*a*) Pictures including certain combinations of persons, such as woman-boy, man-woman, etc., were placed in each battery. (*b*) Fairly equal distribution of pictures judged to be of positive, negative or neutral tone was attempted. (*c*) Attention was given to the amount of activity in the pictures in an effort to obtain a fairly equal distribution. (*d*) Each battery began with a picture considered as likely to be interesting from the subjects' viewpoint. (*e*) Within each battery the pictures were arranged in a sequence that promised variety and interest. Battery *A* was organized to include motion picture sequences 1, 13, 2, 7, 3, 11 spliced in that order for presentation and still pictures in order 6, 8, 14, 4, 12, and 5. Battery *B* sequence for movies was 6, 8, 14, 12, 4, 5 and for stills 1, 13, 2, 7, 3, and 11. These batteries represented a utilization of 12 of the 15 motion pictures developed.

The investigator was interested in the effects of including in Battery *A* two still pictures drawn from motion pictures in the same battery. The hypothesis was that subjects would readily identify the single frames as representing motion pictures they had just seen. This proved to be correct. The battery was given to three subjects from the same grade as the experimental group and all of them spontaneously identified the still pictures in their stories. As a result the two pictures were withdrawn and the final batteries were comprised of the stimuli indicated above.

The picture situations in each battery totalled 12, six in motion and six in still form. It appeared reasonable that six stories based on motion and six on still pictures would elicit a fair sampling of the fantasy behavior of the subjects. The purpose was not to develop so extensive a battery as the Thematic Apperception Test, but only to test the hypotheses of this study, which are basically methodological.

B. Collection of the Data

1. *Selection of the Subjects*

In order to satisfy the criterion that most of the pictures contain someone with whom a subject might identify, boys were used as subjects in the pictures. The population selection for the experimental group comprised all of the boys, 50, in the tenth grade of the Laboratory School, The University of Chicago. The ages of the boys ranged from 14 years, 1 month to 16 years, 3 months. The median age was 15 years, 5 months. The intelligence test scores (Stanford-Binet) ranged from 93 to 159, with a median score of 134.5. The standard deviation was 13.0.

Four different sequential presentations of still and motion pictures were used and the subjects were assigned in the order in which they appeared for testing. The situations were rotated in the following manner.

a. Battery *A*—Motion Picture Series, Still Picture Series.
b. Battery *A*—Still Picture Series, Motion Picture Series.
c. Battery *B*—Motion Picture Series, Still Picture Series.
d. Battery *B*—Still Picture Series, Motion Picture Series.[3]

This procedure was based on administrative convenience but the question may be raised, were the subjects distributed to the presentations in a random manner? The investigator knows of no factors that might invalidate such an assumption. It might be of interest, however, to examine the four group-

[3] These groups are coded in the following pages thus: *IMS*, *ISM*, *BMS*, *BSM*, respectively.

ings of subjects with respect to the two factors, age and intelligence, to determine whether they might be assumed to have been drawn at random from the same population.

The method used to make this evaluation is analysis of variance.[4] We will test the hypothesis that the four groups are random samples drawn from the same population by deriving two independent estimates of the population variance, one based on the variance of the group means and the other on the average variance within the groups. The test of the hypothesis is to determine whether the ratio (F) between these estimates lies below the value in the table of F for a particular level of significance. First we will apply the analysis to the data for age, then IQ. The mean age (months) for each group is *AMS*—181.5, *ASM*—186.5, *BMS*—185.1, *BSM*—184.8.

We may then formulate the hypothesis: the obtained differences among the means are no greater than those likely to occur by chance. The results of the analysis of variance are shown in Table 3. To apply the test of significance we obtain the F ratio as follows: $52.67/28.42 = 1.85$. On entering

TABLE 3
ANALYSIS OF VARIANCE FOR GROUPS BY AGE

Source of variation	*df*	Sum of squares	Variance
Among groups	3	158.01	52.67
Within groups	44	1,250.59	28.42
Total	47	1,408.60	

the table for F (24) with 3 and 44 degrees of freedom we find that the value for rejecting the hypothesis at the 5 per cent level is 2.82. We must, then, accept the hypothesis that the obtained difference could be accounted for by chance fluctuations. In other words, the obtained differences might reasonably be expected on repeated sampling.

The same procedure was employed to evaluate the differences in IQ scores for the four groups of subjects. The mean IQ for each group was *AMS*—138.3, *ASM*—131.0, *BMS*—128.1, *BSM*—132.3. The hypothesis to be tested is: the obtained differences among the means are no greater than those likely to occur by chance. The results of the analysis of variance are shown in Table 4.

To apply the test of significance we would divide the among groups variance by the within groups variance to obtain an F ratio. However, since the within groups variance is the larger, we need not make the computation.

[4] A discussion and formulae for analysis of variance may be found in Lindquist (24).

TABLE 4
ANALYSIS OF VARIANCE FOR GROUPS BY *IQ*

Source of variation	*df*	Sum of squares	Variance
Among groups	3	636.58	212.19
Within groups	44	10,283.90	233.73
Total	47	10,920.48	

The F ratio required for significance at the 5 per cent level for 3 and 44 degrees of freedom is obviously not reached and the hypothesis is accepted. We may conclude that the differences obtained between the *IQ* means for the groups might reasonably be expected on repeated sampling.

It should be pointed out that the evaluation of different group means and of the individual differences between them involves the assumption that the variance within a group is the same except for chance, from group to group (24). We may increase our confidence in the results if, rather than making the assumption, an exact test of the hypothesis of homogeneous variance were applied to the observed data. Following Lindquist (24) we apply the formula $X^2 \frac{6.907\,8n}{3n + k + 1} (n \log_{10} S^2_{\text{w}} - N_i \Sigma \log_{10} S_i^2)$ to the data for age and *IQ* of the groups. The required X^2 value for 3 degrees of freedom at the 5 per cent level is 9.837. Since the obtained X^2 values are 2.978 for age and 4.274 for *IQ*, we are obviously justified in retaining the hypothesis of homogeneous variance.

The conclusion seems warranted, then, that insofar as age and intelligence are concerned there were no significant differences among the four groups. This fact suggests the probability that there are no significant differences in other factors among the groups and makes it possible to interpret any differences in the experimental variables with greater confidence. The importance of these conditions will be considered in greater detail in Section IV because of their bearing upon the assumptions of randomisation and homogeneity of variance underlying the use of analysis of variance techniques.

2. *Administration of the Tests*

The first few minutes of each testing period were devoted to establishing rapport with the subject. He was familiarized with the equipment, especially the wire recorder on which a test run was conducted to try out the reception. Practically all of the subjects had had previous experience with the wire recorder. It was placed in an inconspicuous place, and most of the

subjects paid little or no attention to the microphone when the tests got under way. Each subject viewed the images from the same position, approximately 16 feet from the screen. The examiner sat to the right and rear of the subject. He operated all the equipment alone. At no time was anyone else in the room during the testing periods. Because no writing was done the room was comparatively dark.

After a brief orientation about the general purpose of the investigation, the following instructions were given to each subject:

> This is a story telling test. I have some short motion pictures and some still pictures that I am going to show you, one at a time. For each picture I want you to make up a story. Tell what has happened before, and what is happening now. Say what the people are feeling and thinking and how the story will come out. You can make up any kind of story you please. Take as much time as you need. Each time before I show you a picture I will tell you who is in it. Do you have any questions? All right, here is the first picture.

A significant point at which this investigation differed from the usual directions for obtaining stories about pictures was the verbal structuring given in the instructions. This procedure was considered as more likely to direct fantasies into the area of family relationships than if the directions were more indefinite. The verbal structuring given each stimulus is indicated below.

Battery A—Movies

1. This is about a boy and his mother.
2. Here is a family—father, mother, boy and girl.
3. Here is a boy.
4. The next one is about father and mother.
5. The next one is about a boy and his father and mother.
6. The last one is about a boy and his sister.

Battery A—Stills

1. The first one is about a boy and his father.
2. The next one is about a father and mother.
3. This one is about a boy and his father and sister.
4. The next one is about a boy and his mother.
5. This one is about a boy and his sister.
6. The last one is about a boy and his father and mother.

Battery B—Movies

(Directions were the same as for Battery *A*—Stills.)

Battery B—Stills

(Directions were the same as for Battery *A*—Movies)

The investigator did not press the subjects at any time. Every effort was made to insure that the subject understood what was expected of him, but no effort was made to insure conformity to the instructions. It is common clinical practice to press the subject at certain points to clarify ideas or to elaborate a story that is considered too short. Rapaport (34) discusses the rationale, consequences, and interpretative significance of such practices. To have pressed the subjects in the present study would have introduced another variable, the investigator's inquiry, which might bias the results. It was considered preferable for the examiner to be as much a constant as possible in the experimental situation.

The procedure adopted in this study may not be the most appropriate clinically, but seemed best for the purpose of this research. Undoubtedly more material would have been obtained had the subject been asked for elaboration and clarification of ideas at certain points. The stories were probably less concrete, shorter and sometimes more ambiguous than they might otherwise have been.

3. *Preparation of the Data for Analysis*

All material on the wire recorder was transcribed. Each story was coded to indicate the story number, Battery *A* or *B*, motion picture (*M*) or still picture (*S*) stimulus, and order of presentation, *MS* or *SM*. In clinical practice responses to thematic apperception tests are written by the examiner. When the subject talks at his own pace, the examiner frequently has to slow down the subject in order to get most of the material on paper. This is not necessary when recording apparatus is used.

The examiner was impressed by the hesitations, uncertainties, repetitions of words and phrases that were revealed in most records. His experience with taking records in handwriting leads him to believe that considerable material is lost when a recorder is not used. The question, of course, of the significance of such material remains undetermined. An investigation might well be made to discover the interpretative significance of the differences between recorded and examiner-written responses. The relationship between the amount of material obtained and the time taken to obtain it is examined in Section III.

III. STORY PRODUCTIVITY AND ASSOCIATED RELATIONSHIPS

A. Length Characteristics of the Stories

Since each of the 50 subjects received 12 stimuli, six motion and six still pictures, 600 stories were obtained. Because a verbatim record was procured by using a wire recorder, a word count constituted a precise measure of the length of the stories. Every word was counted except the expression, "ah." A frequency distribution of story lengths is given in Table 5.

The stories ranged in length from 12 to 237 words. The mean word length of the 600 stories was 55.31, standard deviation 28.34. The mean productivity per person was 663.74 words, standard deviation 256.12. The

TABLE 5
Frequency Distribution of Story Lengths

Range and scores	Frequency
226-240	1
211-225	0
196-210	0
181-195	1
166-180	1
151-165	3
136-150	3
121-135	12
106-120	17
91-105	26
76- 90	40
61- 75	94
46- 60	148
31- 45	158
16- 30	88
1- 15	8

mean word length of the 300 stories given in response to motion picture stimuli was 59.02, standard deviation 26.16 while the mean word length of the 300 stories in response to still pictures was 51.60, standard deviation 29.86.

These figures are considerably below the minimum deemed necessary for adequate interpretation by Murray (31). "Stories from a sane adult averaging less than 140 words per story usually indicate lack of rapport and coöperation, lack of self-involvement. As a rule they are not worth scoring." This statement has been challenged by Henry (17), whose Indian records averaged about 200 words per person. That the significance of the factor

of length in the interpretability of stories may be a function of the system of analysis used is suggested by Henry. The significance of the length factor is considered in greater detail in Section IV.

Another interesting characteristic of productivity is the extent to which story lengths in response to motion and still pictures were associated. The product-moment correlation coefficient for the 50 cases was +.81. For Battery *A*—26 cases—it was +.81; for Battery *B*—24 cases—it was +.82. These are significant correlations at the 1 per cent level, which indicates that individuals tended to occupy the same relative positions in the distributions of length for motion and still pictures.

These relationships may point to consistency in the respondent's behavior or to similarities in the kinds of stimuli presented. In this instance the still pictures were those taken from *other* motion pictures and *judged* to be equivalent. This suggests the desirability of determining the relationship between length of response to motion pictures and to their presumed equivalents, that is, the still pictures taken from the motion pictures. In this instance the subjects are different. The comparison is made between the mean productivity of each motion picture and the mean productivity of the equivalent still picture. The number of situations was 12 and the product-moment coefficient of correlation was +.86. This is a statistically significant correlation at the 1 per cent level.

These results seem to indicate that there is both consistency among subjects and similarity in stimulus power with respect to responses length to a motion picture and its equivalent still. The relative values of various stimuli are considered in Section IV.

B. Intelligence and Productivity

The relationship between the amount of material produced in reaction to picture stimuli and the intelligence test scores of the subjects producing stories has not been systematically studied. Statements about such relationships are usually by-products of investigations with different objectives. It seems reasonable to assume that individuals in the lowest end of the intelligence score distribution would produce less material than brighter individuals. Support for this view has been given by Despert and Potter (10). "Dull children yield less information than the more intelligent ones, although there is no absolute relation between *IQ* and Productivity." This does not imply that the story telling type of projective instrument has little value for studying individuals of inferior intelligence. The work of Sarason (36) with mental defectives has shown that the Thematic Apperception Test is a useful clinical tool.

The *IQ* scores and the total productivity scores for the 50 subjects were correlated by the product-moment method. The obtained value was +.043 which is not significant at the 5 per cent level. This result indicates that intelligence test scores on the population employed are not significant indicators of the total length of stories given in response to the pictures.

C. Relationships Involving Time and Productivity

When reacting to a still picture, the subjects had the stimulus before them throughout the period of response. On the other hand, the subjects saw a complete sequence when reacting to the motion picture, before telling a story. Another possible experimental design might have been to limit exposure of the still picture stimuli to the exact time it took for projection of each equivalent motion picture. The investigator considered it desirable to follow the practice used clinically, that is, to allow the subjects to have the stimulus available throughout the story-telling. This is not technically possible with the motion picture.

Data on response time were made available in the following manner. Subjects were "clocked" on still pictures from the time the first stimulus was presented until the story about the last stimulus was completed. Subjects were "clocked" on motion pictures from the time the projector was started for the first stimulus until the story about the last stimulus was completed. The total projection time for Battery *A* stimuli (motion pictures) was 98 seconds, for Battery *B* 83 seconds. These amounts of time were deducted from the total reaction time of each subject before making comparisons.

These data are best viewed as approximations for the purpose of comparing the relative time taken for reactions to motion and still pictures. This is true because total time for each group of stimuli was obtained instead of reaction time to each stimulus, which would have made finer differentiations possible. However, there may be some value in studying the results with these conditions in mind.

The *average* amount of response time for each subject was about 10 minutes. This was fairly evenly distributed between motion and still pictures. In order to determine whether the differences between any pair of means were significant the *t*-test was applied. No difference was significant at the 5 per cent level. It may be concluded that the four groups took about the same amount of average time to react to the motion and still pictures.

That individual differences were operating is evidenced by a study of the relationship between the response times of the 50 subjects to motion and to still pictures. The product-moment coefficient of correlation was +.635.

This value is statistically significant at the 1 per cent level. This may be interpreted to mean that there was a significant tendency for individuals to take about the same amount of time to respond to both types of stimuli. However, there remains a large portion of variance not accounted for.

To many clinicians 10 minutes will seem a rather short period of time for responses to 12 stimuli. While it has been shown above that the stories obtained are somewhat shorter than might be procured under clinical conditions, another important reason for the shorter response time was the use of the wire recorder. Since the mean productivity for the 50 subjects was about 660 words and the mean response time was 10 minutes, a hand-written record would have required a rate of about 66 words a minute, to keep up with the narration. The average *word* from a sampling of the stories contained about five letters. This means that a person recording by longhand would have to write about 330 letters a minute. This is an impossible demand as indicated by studies in handwriting speed. "Nifenecker found 106 letters per minute to be the average rate for 161 sales clerks, billers, checkers and bookkeepers" (26). And these occupations have high requirements for rapid writing. It may be concluded that the relatively large amount of material obtained within the short response time can be attributed largely to the use of the wire recorder.

A significant relationship was found between the total response time and the productivity of the 50 subjects. The product-moment coefficient of correlation was +.46 which is significant at the 1 per cent level. However, there remains a considerable portion of variance not accounted for. We may conclude that there was a slight tendency for those who told longer stories to take more time to do it.

Data were collected which made it possible to determine the relationship between the time taken for projection of the motion picture stimulus and productivity elicited by that stimulus. The projection time for the 12 motion pictures ranged from 11 to 21 seconds. The mean time was 15 seconds. The productivity score for each stimulus was taken as the sum of all the length scores for the 24 subjects reacting to that stimulus. The product-moment coefficient of correlation between these variables was +.245. This value is not significant at the 5 per cent level. The scatter diagram of these scores did not indicate any non-rectilinear relationship between the time of projection and the productivity for the stimuli.

D. Summary

The purpose of this section was to report evidence bearing on the productivity aspects of the stories. The evidence seems to support the following conclusions.

1. The lengths of the stories produced in this research probably compares favorably with those reported in several other investigations.

2. There was a considerable tendency for individuals to be productive in response to motion and still pictures to the same degree ($r = +.81$).

3. The intelligence test scores of the tested population are not a significant indication of the amount of material given in response to the pictures ($r = +.04$).

4. The large amount of material obtained within the short response time may be attributed to the use of the wire recorder.

5. A tendency was noted for individuals to take about the same amount of time for responding to six motion and six still pictures ($r = +.64$).

6. There was a moderate tendency for those who gave more materials to take more time to do it ($r = +.46$).

7. There was no significant relationship between amount of time for projection of a motion picture and the amount of material elicited by that picture ($r = +.25$).

IV. METHODOLOGICAL ASPECTS OF THE STUDY

A. Introduction

In this section we will discuss the bearing of the data upon the following major methodological questions.

1. Are there significant differences between the motion and still picture methods of presenting picture situations?
2. Are there significant differences among the picture situations in power to produce meaningful material.
3. Are there significant differences in the results when motion pictures are presented *before* and *after* still pictures?
4. Are there significant differences between the two batteries of tests?
5. Can evidence be obtained of the stability of this experiment?

The answer to these questions required some quantification of the basic "story" data.

The stories were scored in terms of four separate procedures. The chief objective in using four methods was to obtain a measure of the consistency with which certain hypotheses were tested by the data. The scoring methods are called need-press, intraceptive language, discomfort-relief words, and rating scale.[5] Each procedure is described below. Attention is called to its rationale and previous uses, if any. A report on scorer reliability is given for each method.

Following the presentation of scoring methods, the appropriateness of analysis of variance as the statistical technique for treating the data is indicated. This technique is then applied to the data and the results are presented and discussed. A synthesis of the results of the analyses of variance is attempted to determine whether consistency among the scoring methods was achieved.

B. Scoring Methods

1. *Need-Press Scoring*

The need-press analysis used in this study was developed by Murray (29) and his co-workers at the Harvard Psychological Clinic. Murray defines in great detail the conceptual framework within which "need-press" as psychological variables are rooted. This scheme of analysis has been used by many investigators including Sanford, Wells, and Deabler (35, 42, 9).

[5] A recent summary of methods of scoring and analysis used by leading workers in the thematic apperception field has been made by Wyatt (11).

For the present study the definitions of needs and press formulated by Murray (30) were followed. He describes the general conception of need-press as follows:

> The scheme is based upon a conception of the mutual dependence and interaction of forces—forces within the personality and forces originating in things or people outside the personality. The forces engendered within the personality have been termed *needs* (or drives) and the social or inanimate forces with which the subject has intercourse have been termed press (plural press). It seems that almost any behavioral event is susceptible of analysis into needs and press (30).

Each story was analyzed in terms of 35 needs and 22 press.

The first analysis was carried on by the investigator. Each story was separately attached to a 5 x 8 card and the cards were shuffled. Thus the analyst did not know the identity of the subject or whether the story was a response to a motion or still picture. A need or press was tallied each time it was identified in a story. The rating scale for strength of the variables, which is used by some workers using this method, was not used. However, one measure of strength, frequency of occurrence, was obtained. In this fashion each story was assigned a score representing the sum of the needs and press identified.

To get a measure of investigator reliability, a second analysis of the data was made more than three months after the first analysis. The procedure described above was repeated. A sampling of 230 stories was used. The scores obtained from the two analyses were correlated. The product-moment coefficient of correlation was +.94. It represents the relationship between scores given independently by the same person to the stories. The coefficient obtained does not indicate the extent of agreement in identifying the same needs and press in both analyses. A separate tally revealed that out of 1,132 items there was agreement on 657, or 58 per cent.

Tomkins (39) indicates in his discussion of interpreter reliability the problems involved in obtaining agreement in analyzing test protocols. He points out that reliability coefficients of correlation ranging from +.30 to +.96 have been reported in the literature. "At the moment each investigator is a law unto himself." This conclusion at least suggests that there is a long way to go to reduce interpreter variability in analyzing data of this kind.

In the light of Tomkins' review it seems reasonable to conclude that the interpreter reliability obtained in this study is superior to much of the work that has been done.

2. *Intraceptive Language Scoring*

Another approach to scoring the stories involves an analysis of the "intraceptive language" used by the subjects. This does not comprise a system of analysis in the same sense as "need-press"; it is merely one aspect of form analysis as used by investigators in this field.

The definition of intraceptive language will become clearer if it is considered within the framework of descriptive and interpretative language. All comments or remarks that go beyond a routine description of the stimulus may be called interpretative. For example, the remark "the father is talking to the mother," is considered descriptive of the stimulus, while the remark, "the father is explaining something to the mother" is interpretative. The latter remark adds something imaginative to the first remark. All intraceptive remarks are interpretative by definition but they add something by revealing individual needs, desires, motives, feelings, attitudes. Intraceptive language is more personalized language than other interpretative remarks. It tells more about the private world of the teller. "He is waiting there" is interpretative but "he is waiting there impatiently" is, in addition, intraceptive. "He is sad," "he left disgustedly," "he is afraid to confide" are other examples of intraceptive remarks.

This category of analysis was used quantitatively by Henry (17) as one indicator of "inner life freedom." The intraceptive language was most frequently identified as a phrase which contained the appropriate feeling elements. In the analysis each story was attached to a 5 x 8 card and all the cards were shuffled. Thus, the investigator did not know the identity of the subject or whether the story was a response to a motion or still picture. This procedure resulted in a score for each story.

In order to obtain a measure of the consistency with which the investigator identified intraceptive language, a second analysis was made more than three months after the first analysis. The same conditions were observed in analyzing a sampling of 312 stories. The scores obtained from the two analyses of 312 stories were correlated. The product-moment coefficient of correlation was +.81. From a total of 556 items, 359, or 65 per cent, were identical in both analyses. These indices of reliability were judged to be acceptable for the purposes of the study. The stories tended to receive scores of the same magnitude on repeated analysis.

3. *Discomfort-Relief Word Scoring*

The basic procedure in the "Discomfort-relief" method of scoring the stories is to identify words which indicate discomfort (suffering, tension,

pain, unhappiness) and relief (comfort, satisfaction, enjoyment). The procedure was developed by Dollard and Mowrer (11) to measure tension in written documents. It has not been applied in the field of thematic apperception study although these authors suggest the possibility. The psychological rationale for this approach to scoring is given below.

> Possibly the psychological transaction involved in D.R.Q. scoring can be indicated as follows: The scorer rehearses the sentence. As he does so, the sentence or thought-unit produces tensing or relaxing responses in the scorer, or he experiences no changes in tension level. The scorers are then required to make a verbal response to the cue produced by the tension change. This verbal response is the score. He says "drive" in case tension is increased; "reward" in case tension is reduced and "zero" if rehearsal has produced no change in tension level. Usually this rehearsal transaction occurs so rapidly that it is described by the scorer as "intuitive."
>
> The reliability of judgment concerning the D.R.Q. gives evidence of the fact that the rise and fall of human tensions are well registered in common speech and that likewise this speech is sufficiently widely disseminated so that a considerable number of people can make accurate judgments in respect to tension movement (11).

The method is most closely related to attempts in the thematic apperception field to identify emotionalized words of a specific type such as "aggressive" (4). A preliminary attempt was made to use "emotionalized words" as a scoring technique. They were defined as words which express personal feelings, without differentiation as to part of speech. One hundred fifty-six stories were analyzed in this manner, the total number of such words constituting the score for each story. One month later the same stories were analyzed by the discomfort-relief method described above. The first 100 words identified by each method were then compared and it was found that 90 were identical. This made it evident that either one of these methods could be used. Very little would be gained by using methods which gave similar results. Because of the existing reliability studies on the discomfort-relief word approach the decision was made to use it instead of "emotionalized words."

The discomfort and relief words were identified in all of the stories. Again the identity of the subject and whether the story was a response to a motion or still picture were not known. The sum of the discomfort and relief words constituted the score for each story.

A second analysis using 312 stories was performed by the experimenter under the same conditions as those for the first analysis. The scores, story by

story, for these two analyses were correlated. The product-moment coefficient was +.90.

Two independent judges then performed the discomfort-relief word analysis using the above mentioned article for directions. These judges were college graduates without special psychological training. The scores obtained by the investigator for 312 stories were then compared with those obtained by each scorer. The product-moment coefficients were +.83 and +.75. This correspondence in attributing scores to particular stories does not imply that the scores were derived by identification of the same words by each scorer. It is apparent, however, that the three scorers showed considerable agreement in attributing the same total *scores* to particular stories. This accomplishment satisfies the purpose for which these analyses were made—to differentiate among stories.

4. *Rating Scale Scoring*

The quantitative treatments of the data which have been described—need-press, intraception language, and discomfort-relief words,—may be considered analytical approaches in the sense that they involve detailed analysis of a story into component elements and variables which were then added to give a score for the story. The rating scale method to be described here is a whole-story approach. It is based on the assumption that a story which is descriptive only will tell very little about the teller's private world, while a story that is projective will reveal much more about the teller's private world. The definitions of "descriptive" and "projective" stories will be made clearer by examining the procedure which was employed.

All of the stories (600) had been placed on 5 x 8 cards, in order that the judge would not know the identity of the subject or whether the story was given in response to a motion or still picture. It is apparent from Table 5 that there is considerable range in the length of stories. More than half of the stories (306) fall within the interval 31-60 words. On the assumption that the length of the stories should be controlled when the rating was done, the subject responses were distributed into five groupings: 1-30, 31-60, 61-90, 91-120, and 121 up.

Each of the five groups of stories was judged separately. The following instructions were used by the investigator on two different occasions and once by an independent judge.

> The purpose of this sorting procedure is to locate each story on a four-point scale. The scale indicates degree of self or ego involvement of the story teller. The scale might also be thought of as a continuum

along which each story could be placed, with complete descriptiveness (non self-involvement) at one end and maximum projectiveness (self-involvement) at the other.

Description in this context means that the story simply relates the physical aspects of the stimulus or tells about events in words which imply no motives, desires, attitudes, feelings.

Projection implies the presence of words or statements involving feelings, attitudes, desires, motives.

Read carefully the definitions of each category before you begin the sorting procedure.

Category I

The story is entirely descriptive. Words describing feelings, motives, desires, attitudes are not used.

Category II

A considerable portion of the story is descriptive. Some words implying feelings and attitudes are used.

Category III

The story includes a considerable amount of language involving feelings, attitudes, desires. Some descriptive words are used.

Category IV

Statements implying feelings, attitudes, desires are the dominant feature of the story.

1. Read each story completely and separately. Then place it in one of the four categories before you. If you cannot readily classify the story into one of the four described categories, place it in a separate pile called "not readily classifiable."
2. After you have performed the above operation you will have a clearer definition in your mind of the relative meanings of the categories.
3. Then go to Category I and decide whether each story should remain in that category or should be shifted to another category. Make any changes you think necessary.
4. Then go to Category II. Carry on the same operation as in (3).
5. Then go to Category IV. Carry on the same operation as in (3).
6. Then go to Category III. Carry on the same operation as in (3).
7. After the above operations have been performed you will have a better basis for re-examining the stories which you placed in the "not readily classifiable" category. Now go through these stories and put each one into one of the four categories described.

This procedure resulted in the assignment of a score to each story, that is, the four points on the continuum were treated as scores.

In order to obtain a measure of reliability of the experimenter's use of the rating scale method, the sorting of the 600 stories was repeated three months later. The same procedure was followed. An objectivity measure

of this procedure was obtained by use of an independent judge. He sorted the 306 stories in the 31-60 word length grouping. If we consider complete agreement and one-class difference as representing substantial agreement in placing a story in a given category on repeated sortings, the extent of agreement was 98 per cent for the investigator with himself and 97 per cent for the investigator with an independent judge. The results of these procedures indicate that reasonable success was achieved in differentiating stories on the scale.

C. The Significance of Analysis of Variance and Factorial Design for This Study

It has been shown how quantitative measures were assigned to each story by four different approaches: need-press, intraceptive language, discomfort-relief words, and rating scale.

We may get some understanding of the factorial design from Table 6. Combinations of the various cells represent the factors to be studied.

TABLE 6
Experimental Design

Battery *A*		Battery *B*	
Movies	Stills	Movies	Stills
A	*B*	*C*	*D*
E	*F*	*G*	*H*

The basic definitions are as follows:

1. Each of the above cells represents 12 subjects and 6 picture situations. For example, *A* represents 12 subjects and 6 motion picture situations. The cell then includes scores for 12 × 6 or 72 stories. *B*, then, would include Situations 7-8-9-10-11-12 for the same subjects.
2. Battery *A* refers to $A + B + E + F$ and Battery *B* refers to $C + D + G + H$.
3. Block 1 includes $A + B + G + H$ and Block 2 includes $E + F + C + D$.
4. Methods refers to $A + E + C + G$ for motion pictures and $B + F + D + H$ for still pictures.
5. Order of presentation refers to $A + F + C + H$ for first order of presentation and $E + B + G + D$ refers to second order of presentation.
6. Picture situations includes the combination of each of the 1-12 situations in *AB* and *EF* with their counterparts in *CD* and *GH*.

This introduces the factors in the design of the study which are considered. The kinds of comparisons to be made have been suggested by examination of the experimental design.

1. Are there any differences between the methods?
2. Are there any differences among the picture situations?
3. Are there any differences between orders of presentations?
4. Are there any differences between the batteries?
5. Are there any differences between the blocks?

A common approach to answer such questions would be to re-state them as null hypotheses;[6] then test the significance of the differences between the means for each comparison by using the *t*-test.[7] Simple direct comparison of differences in means did not appear to be the appropriate technique for analyzing the apparent variability in the data. It seemed that the factors—subjects, methods, picture situations, orders of presentation and batteries—were simultaneously operating in the experimental situations. Therefore, a method which would make possible tests of significance of these several factors singly and in *their various combinations* would be more appropriate. Such a method is available in the use of analysis of variance in factorial design.[8]

Analysis of variance is most closely associated with the work of Fisher (14) in the field of agricultural experimentation. In recent times, efforts have been made to apply these methods to educational and psychological research, illustrated by Lindquist (24) and Johnson (20). Analysis of variance has been well defined by Snedecor (37).

> It is a technique for segregating from comparable groups of data the variation traceable to specified sources. It furnishes an estimate of experimental error freed of that part of the variability whose origin is known. In conjunction with an associated test of significance, it affords a basis of judgment as to whether or not several groups are samples from a single homogeneous population. Through the use of degrees of freedom it is equally well adapted to the treatment of large or small samples. It lends itself readily to the design of efficient experimental procedures, and to tests of experimental technique.

A design can be called factorial when it involves two or more factors or

---

[6]The generalized statement of the null hypothesis is to formulate the hypothesis that some particular population value is zero. An example here would be "the difference between the means of the motion and still picture methods are not significally different from zero." For a discussion of this procedure see Walker (40).

[7]A discussion of this procedure can be found in Garrett (16).

[8]The point raised here is still somewhat controversial. It is concerned with two methods of experimental design: one, to control all factors while manipulating a single variable and two, to permit various factors to operate simultaneously within a design that admits of measurements of those factors through statistical method. For discussion of the relative merits of these arguments, see Fisher (13), Crutchfield (6), Peters and Van Voorhees (33).

variables in all of their combinations (2). Analysis of variance and factorial design were adopted for this study because this method promised the most precise answers to hypotheses being tested and the most efficient use of the data available. Concerning an advantage of factorial design over the single factor experiment Fisher (13) has remarked, "at the same time as it is made more comprehensive, may also be made more efficient, if by more efficient we mean that more knowledge and a higher degree of precision are obtainable by the same number of observations."

Before examining the data five aspects of factorial design should be discussed.

1. *Experimental Error*

Baxter (2) has given a good description of experimental error.

> It is the residual variance or that variance about the grand mean which cannot be accounted for by the factors or controls of the experiment. The experimenter designs his problems so that portions of the total variance may be said to be the result of certain known factors. . . . The unassignable portion is the basis for tests of significance: it is analogous to the standard deviation from which, in the older forms of statistics, the standard error of the difference might be calculated. An efficient design seeks to get a precise estimate of the experimental error so that all the variables will have a sound basis for evaluation. An essential condition of properly designed experiment is that it must provide its own estimate of experimental error.

The experimental error, then, consists of the residual variance which becomes the basis for testing the significance of factors whose effects are being studied. Individual differences among the subjects with respect to these variables are probably one source of variance in the data. The identification of individual differences is not a purpose of this study, which is primarily focussed on studying the influence of two methods, motion and still pictures. For a given purpose, a known source of variance may be included in experimental error if its effects can be randomised. Therefore, the variance which may be due to individual differences is put into experimental error.

2. *Randomisation*

The procedure for assigning individuals to the various experimental situations (treatments) was discussed in Section II. The probability that randomisation was achieved was also considered. It probably would have been better to assign individuals to the four experimental conditions on a strictly random basis, for randomisation improves the quality of obtained evidence by assuring that an estimate of experimental error is unbiased.

3. *Replication*

A replication refers to that unit within which all treatments (experimental situations) can be found. In other words, the experiment is, in effect, repeated. Within this framework a replication is achieved when both batteries, both orders of presentation, both methods, and all picture situations are included in one unit. This, in fact, is a definition of a "block" as used hereafter. In essence, it means that the data have been so arranged for analysis that two repetitions of the experiment may be said to be achieved. This element of the design increases our confidence in the results. The nonsignificance of block differences would reflect stability in the experiment.

4. *Heterogeneity*

In agricultural work the problem of soil heterogeneity had been an obstacle to the precise estimate of experimental error. The development of ingenious kinds of experimental design have helped research people in this area to solve this problem.[9] In psychological experiments heterogeneity is more complex, for it is found in the multifarious factors which characterize individuals. Soil heterogeneity in agriculture is related to the spatial aspects of a block. Individuals are dynamic, changing in time, and differing with respect to most characteristics. The experimenter, then, must be sensitive to the relationships of any of these characteristics to the particular variables he is investigating; otherwise, heterogeneity will be increased and his results will be less precise.

Within the context of this investigation, no attempt is made to control heterogeneity within the individual subjects. An attempt is made to randomize its effects on the estimate of experimental error. An experimental design could be devised which would identify and estimate the variance which is attributable to individual differences but that implies different questions than set forth for study in this research. But it does suggest that flexibility is possible in the kind of experimental design that is amenable to analysis of variance techniques.

5. *Interaction or Discrepance*

It has been shown that an advantage of factorial design is that it provides information, not only about single factors (main effects), but also about the combinations of these factors. As an example, let us take batteries and methods. If the methods means are significantly different, does this difference

[9]For a detailed discussion of the relationship of heterogeneity to factorial design see Fisher (13).

obtain for both batteries? We must look to the interaction of battery and method to obtain an answer. If this interaction "*F*" ratio is non-significant, the method means may be said to be significantly different independent of the batteries involved. If this interaction "*F*" ratio is significant, there is said to be interaction of the two variables.

This interaction between two variables must be accounted for. Kendall (21) has warned that interaction ratios may be significant without the presence of real interaction between factors. Whether an interaction is a real effect or not can be answered only by resort to the data, the design and the experimenter's knowledge of the total situation.

D. Application of Analysis of Variance to the Data

The null hypotheses to be tested by the application of analysis of variance to the data are stated as follows:

> There are no significant differences between the means of students' scores on motion and still pictures (methods).
>
> There are no significant differences among the means of students' scores on the 12 picture situations (situations).
>
> There are no significant differences between the means of students' scores in Batteries *A* and *B* (battery).
>
> There are no significant differences between the means of students' scores on the groups of six motion and six still pictures, whether presented first or second (order).
>
> There are no significant differences between the means of students' scores on the pictures in the two Blocks (blocks).
>
> All interactions rows represent null hypotheses about the variations in the rows.

The results for each of the four analyses of variance[10] will be presented and briefly discussed in the order: need-press, *intraceptive* language, discomfort-relief words, and rating scale. Then the results of the four approaches will be summarized and discussed in greater detail. The same procedure will be followed for each presentation.

The first step in the analysis was to examine all of the interactions for significance. Johnson and Tsao (20) have pointed out that:

> The policy is rather often followed, on the basis of findings in agricultural experiments, to assume that higher-order interactions are not significant. It is recommended here that interactions of all orders should be tested when the numerical measure of the interaction can be obtained from the data.

[10]The formulae used for computation of analysis of variance may be found in a paper by Johnson and Tsao (20). Also included are the mathematical derivations.

All of the interactions which had "*F*" ratios significant at least at the 5 per cent level were identified. Secondly, all interaction values which did not meet these standards were combined with the residual sum of squares to comprise the experimental error. The experimental error mean square (obtained by dividing the residual sum of squares by the appropriate degrees of freedom) then became the denominator for obtaining the "*F*" ratios for all of the main effects and the interactions previously identified as significant (1 per cent level), or doubtfully significant (5 per cent level). The results of this procedure are shown in the summary tables. These tables will provide the bases for presentation and discussion of results.

The procedure in presenting the results will be as follows: the tables will be presented and the significant main effects and interactions will be brought up in this order—methods and interactions involving methods, situations and interactions involving situations, the main effects, order, battery and blocks.

1. *Need-Press Results*

We see from Table 7 that the hypothesis that there is no significant difference between the methods means is rejected at the 1 per cent level. This does not tell us which method mean is greater. For this we go to the quantified data[11] and find that it is the motion picture method mean which is

TABLE 7
SUMMARY OF ANALYSIS OF VARIANCE FOR NEED-PRESS VALUES

Source of variation	*df*	*ss*	*ms*	*F*	Hypothesis*
Order	1	4.87	4.87	1.60	*A*
Blocks	1	4.87	4.87	1.60	*A*
Situations	11	427.10	38.83	12.73	*R*
Method	1	35.50	35.50	11.64	*R*
Battery	1	4.17	4.17	1.37	*A*
Order x Blocks	1	47.28	47.28	15.50	*R*
Order x Method	1	29.80	29.80	9.77	*R*
Blocks x Battery	1	29.80	29.80	9.77	*R*
Situations x Battery	11	76.10	6.92	2.27	*D*
Method x Battery	1	47.26	47.26	15.50	*R*
Order x Blocks x Battery	1	35.49	35.49	11.64	*R*
Error	544	1,660.84	3.05		
Total	575	2,403.08			

*R represents rejection of hypothesis; A represents acceptance of hypothesis; D indicates doubtful region.

[11]The "quantified data" is available in the unpublished Ph.D. dissertation on file at the University of Chicago.

significantly greater than the still picture method mean. But the answer is not clear cut for the interactions indicate that the methods variance is entangled with those of battery and order.

In order to determine what is involved in these interactions, we take our analysis one step further. Let us take the interaction method battery. We may surmise that the superiority of the motion picture over the still picture method is attributable to the results in one battery. We may test this hypothesis as follows:

$$\text{Battery } A \qquad \frac{(492)^2+(445)^2}{144} - \frac{(995)^2}{288} = 0.43$$

$$\text{Battery } B \qquad \frac{(599)^2+(503)^2}{144} - \frac{(1044)^2}{288} = 82.35$$

These squares of differences in means between methods for each battery may then be evaluated by the error variance of 3.05.

$$\text{Battery } A \quad 0.43/3.05 = 0.141$$
$$\text{Battery } B \quad 82.35/3.05 = 27.00$$

The "*F*" ratio required for significance at the 1 per cent level for the appropriate degrees of freedom is 6.70. This evidence indicates that the superiority of the method mean for motion pictures may be attributed to its function in Battery *B* rather than in both batteries.

The other instance of significant interaction involving method is the method x order interaction. Inspection of the data reveals that within Battery *B*, where the motion picture method seems to be superior, one-half of the subjects received the motion picture first and the other half received the motion pictures second. Both groups of subjects received higher scores on motion than on still pictures. The method and order of presentation were interacting. This suggests that the method of the motion picture was better whether presented first or second. In Battery *A* the apparent differences among the group scores involving methods or orders of presentation were very small.

The hypothesis that there are no significant differences among the means of the 12 picture situations is rejected at the 1 per cent level. This may be interpreted to mean that the various picture situations differ significantly in their power to stimulate fantasy. There are no significant interactions involving situations, although the Battery x Situations interaction is in the doubtful region, that is, between the 5 and 1 per cent levels of significance.

It is therefore probable that the picture situations means differ significantly independently of other factors in the experimental situation.

When the overall test of significance, the "*F*" test, is significant and there are more than two means involved, we are justified in evaluating the differences between the various means by the *t*-test (24). This procedure is based upon the assumption of homogeneity of variance. The standard error of the mean which is used to test the difference between any two means is taken as the "within groups" or error variance. This can be done most confidently when the above assumption can be made. However, it is possible to apply an exact test of the hypothesis of homogeneous variance to the observed data. This was done following Lindquist (24). The formula used is

$$X^2 = \frac{6.9078\ n}{3n + k + 1}\ (n \log_{10} S'2 - Ni \sum \log_{10} Si^2)$$

where $n = 576$, the number of scores
$ni = 48$, the number of scores in a group or class
$k = 12$, the number of groups or classes

The X^2 value obtained was 14.28 which is less than the 24.72 value required for significance at the 1 per cent level. The degrees of freedom are $k - 1$ or 11. The assumption of homogeneity of variance can be made with confidence and the *t*-test applied to test any pair of means.

Table 8 presents the mean need-press values for the 12 picture situations. The method of evaluating the difference between any two means is as follows, after Lindquist (24).

TABLE 8
MEAN NEED-PRESS VALUES FOR PICTURE SITUATION

Picture situation	Mean
1	5.10
2	3.31
3	2.60
4	2.10
5	3.52
6	2.88
7	4.40
8	3.19
9	3.27
10	3.23
11	4.88
12	4.00

> The standard error of any mean may be estimated by dividing the error variance by the number of cases on which that mean is based and extracting the square root of the result. . . If all groups are of the same size, the estimated standard error of any difference is 1.414 times the standard error of any mean. The number of degrees of freedom for *t* in any test involving these estimated standard errors is the same as the d.f. for the error variance.

Hence,

Error variance = 3.05

SE difference = 1.414 $\sqrt{3.05/48} = .356$

Since the groups are of the same size, the minimum value of any difference at the 1 per cent level of significance may be calculated:

$$2.58 \times .356 = .919$$

This value may be used to evaluate the difference between any pair of means. Differentiations among the picture situation means will be made later in this section when attempt will be made to determine the consistency of the various scoring methods.

The evaluation of differences between the mean stimulus values of the various picture situations is based upon the assumption of independence among the situations. This is not strictly true because it is likely that there were overlapping stimulations from one picture to another. It is recognized also that the failure of a subject to react to a particular picture may have much interpretative significance, such as an indication of emotional blocking. However, if a particular picture is *consistently* low in its power to evoke fantasy material the contention is that it is inferior to a picture that rather consistently evokes interpretable material.

An experimental design that would yield more precise determinations of stimulus values would be one in which random rotation of stimuli could be allowed for.

The hypothesis that the difference between the means of orders of presentation is not significant is accepted (see Table 7). For purposes of interpretation it should be kept in mind that order of presentation refers to the factor in the experimental design that a series of six motion pictures was presented *before* a series of six still pictures to one group of subjects, and *after* a series of still pictures to another group of subjects. As a main effect this factor was not significant. However, order was involved in the significant Method x Order interaction, which has already been discussed.

The hypothesis that the difference between the means of Batteries *A* and

B is not significant is accepted (see Table 7). This may be interpreted as evidence of homogeneity in the data. The group of 24 subjects comprising Battery *A* and the group of 24 in Battery *B* are homogeneous with respect to the variable under consideration.

The hypothesis that the difference between the means of the blocks is not significant is accepted (see Table 7). A block is an arrangement of the data which produced the effect of a replicated experiment. This finding points further to the consistency in the data. It may also be interpreted as a measure of the stability of the experiment.

The previous discussion has accounted for all of the measures in Table 7, except the second order interaction of Order x Blocks x Battery and the first order interactions Order x Blocks and Blocks x Battery. It is quite probable that the assumption of no real interaction between the factors of order, blocks, and battery is tenable, since they represent structural features of the design rather than real interacting variables such as methods and picture situations. The question of interpretative significance of interactions will be considered in greater detail under Synthesis of Results of Analysis of Variance.

2. *Intraceptive Language Results*

As in the previous analysis the methods means are significantly different (see Table 9). Inspection of the quantified data indicates that the difference is in favor of the motion picture. In this instance the method does not interact significantly with other factors because the motion picture method means are significantly greater than the still picture means for both batteries.

TABLE 9

SUMMARY OF ANALYSIS OF VARIANCE FOR INTRACEPTIVE LANGUAGE VALUES

Source of variation	*df*	*ss*	*ms*	*F*	Hypothesis*
Order	1	2.78	2.78	1.31	*A*
Blocks	1	0.44	0.44	——	*A*
Situations	11	220.10	20.01	9.39	*R*
Method	1	25.84	25.84	12.13	*R*
Battery	1	0.06	0.06	——	*A*
Situations x Method	11	39.70	3.61	1.69	*A*
Situations x Battery	11	65.48	5.95	2.79	*R*
Order x Blocks x Situation	11	65.43	5.95	2.79	*R*
Order x Blocks x Battery	1	25.84	25.84	12.13	*R*
Error	526	1,118.55	2.13		
Total	575	1,564.22			

*R represents rejection of hypothesis; A represents acceptance of hypothesis; D indicates doubtful region.

TABLE 10
MEAN INTRACEPTIVE-LANGUAGE VALUES FOR PICTURE SITUATIONS

Picture situation	Mean
1	2.81
2	1.06
3	2.17
4	0.67
5	1.96
6	1.08
7	1.85
8	1.35
9	1.60
10	1.35
11	2.67
12	1.75

The differences among the picture situation means are significantly different. The mean intraceptive language values for the 12 picture situations are presented in Table 10. Before determining the value required for testing differences between means we test the hypothesis of homogeneous variance by the same formula given on page 246. The X^2 value obtained is 15.65, which is less than the 24.72 value required for significance at the 1 per cent level. The hypothesis of homogeneous variance may then be accepted and we proceed as follows:

Error variance = 2.13

hence, SE difference = 1.414 $\sqrt{2.13/48} = .297$.

The minimum value for testing the significance of any difference between means at the 1 per cent level is

$$2.58 \times .297 = .766.$$

The significance of the interaction of Situations x Battery may indicate that the more effective stimuli may be found in both Batteries. Not all of the pictures in one battery were superior to all of the pictures in the other battery.

Order of presentation again was not significant as a main effect.

The Battery and Blocks means are not significantly different, again pointing to the relative homogeneity of the groupings of subjects and the stability of the experiment.

3. *Discomfort-Relief Word Results*

A review of Table 11 indicates that the only hypothesis rejected at the 1 per cent level was that the means among the picture situations are not

significantly different. The mean discomfort-relief values for the twelve picture situations are given in Table 12.

TABLE 11
SUMMARY OF ANALYSIS OF VARIANCE FOR DISCOMFORT RELIEF WORD VALUES

Source of variation	*df*	*ss*	*ms*	*F*	Hypothesis*
Order	1	9.00	9.00	2.36	*A*
Blocks	1	3.68	3.68	——	*A*
Situations	11	460.82	41.89	10.97	*R*
Method	1	5.45	5.45	1.43	*A*
Battery	1	3.07	3.07	——	*A*
Order x Blocks	1	13.44	13.44	3.52	*A*
Order x Method	1	19.50	19.50	5.10	*D*
Blocks x Battery	1	19.50	19.50	5.10	*D*
Situations x Method	11	90.42	8.22	2.15	*D*
Situations x Battery	11	92.80	8.44	2.21	*D*
Method x Battery	1	13.43	13.43	3.52	*A*
Order x Blocks x Situations	11	85.77	7.80	2.04	*D*
Error	523	1,996.56	3.82		
Total	575	2,813.44			

*R represents rejection of hypothesis; A represents acceptance of hypothesis; D indicates doubtful region.

TABLE 12
MEAN DISCOMFORT-RELIEF WORD VALUES FOR PICTURE SITUATIONS

Picture situation	Mean
1	4.48
2	2.10
3	1.90
4	0.73
5	2.79
6	1.77
7	3.25
8	2.00
9	2.08
10	2.50
11	3.17
12	2.60

The test of the hypothesis of homogeneous variance was made, resulting in an obtained X^2 value of 50.82. Since the X^2 value for significance at the 1 per cent level is 24.72, the hypothesis must be rejected. There is evidence, then, to believe that the data are heterogeneous, that is, the assumption of homogeneity of variance cannot reasonably be made.

The minimum value required for testing the significance of the difference between any means at the 1 per cent level would be obtained as follows:

Error variance = 3.82
hence, *SE* Difference = 1.414 $\sqrt{3.82/48} = .399$
then $2.58 \times .399 = 1.03$.

It is questionable, however, whether we would be justified in testing the difference between a pair of means since the value 1.03 was derived on the assumption that the variance is the same within the group of scores for each situation. The importance of this finding will be considered again when the question of consistency among the methods of scoring is taken up.

4. *Rating Scale Results*

The methods means are shown to be significantly different at the 1 per cent level (see Table 13). Inspection of the quantified data reveals this difference to be in favor of the motion picture method. We note however that method is significantly interacting with Battery. We then apply the test to see whether this difference can be attributed to differences in Battery as indicated under Need-Press Results.

TABLE 13
SUMMARY OF ANALYSIS OF VARIANCE FOR RATING SCALE VALUES

Source of variation	*df*	*ss*	*ms*	*F*	*Hypothesis**
Order	1	0.17	0.17	——	*A*
Blocks	1	0.00	0.00	——	*A*
Situations	11	145.63	13.24	26.22	*R*
Method	1	4.34	4.34	8.59	*R*
Battery	1	0.06	0.06	——	*A*
Order x Blocks	1	7.12	7.12	14.10	*R*
Order x Battery	1	3.37	3.37	6.67	*D*
Blocks x Method	1	3.37	3.37	6.67	*D*
Situations x Method	11	14.75	1.34	2.65	*R*
Situation x Battery	11	19.03	1.73	3.43	*R*
Method x Battery	1	7.11	7.11	14.08	*R*
Order x Blocks x Situation	11	21.46	1.95	3.86	*R*
Order x Blocks x Battery	1	4.32	4.32	8.55	*R*
Order x Method x Situation	11	12.38	1.13	2.24	*D*
Blocks x Situation x Battery	11	12.37	1.12	2.22	*D*
Error	500	252.74	0.505		
Total	575	508.22			

**R* represents rejection of hypothesis; *A* represents acceptance of hypothesis; *D* indicates doubtful region.

$$\text{Battery } A \qquad \frac{(363)^2+(370)^2}{144} - \frac{(733)^2}{288} = 0.17$$

$$\text{Battery } B \qquad \frac{(398)^2+(341)^2}{144} - \frac{(739)^2}{288} = 11.28$$

Since the error variance of .505 is the comparison term we obtain:

Battery A $0.17/.505 = 0.34$
Battery B $11.28/.505 = 22.34$

Since the "F" ratio required for significance at the 1 per cent level is 6.70, we may conclude that the superiority of the motion picture method is attributed to its function in Battery B, rather than in both batteries.

Method and Situation interaction is significant. A tentative explanation may be that some situations are more stimulating as motion, while others are more stimulating as still pictures. As in all of the other analyses the situa-

TABLE 14
MEAN RATING SCALE VALUES FOR PICTURE SITUATIONS

Picture situation	Mean
1	3.58
2	2.13
3	2.67
4	1.69
5	2.60
6	2.00
7	2.77
8	2.48
9	2.44
10	2.31
11	3.31
12	2.69

tion means are significantly different. The mean rating scale values for the twelve situations are presented in Table 14.

The test of homogeneity of variance was made and the obtained value of 24.03 falls short of the required value of 24.72 for the 1 per cent level. We may, then, accept the hypothesis with reasonable confidence that the assumption of homogeneity of variance can be made. As before we then determine the minimum difference between means required for significance at the 1 per cent level.

Error variance = .505
hence, SE Difference = 1.414 $\qquad \sqrt{.505/48} = .144$
then $2.58 \times .144 = .372$

The interaction of Situations with Battery may indicate that some of the better pictures are to be found in both Batteries. That is, not all of the pictures in one Battery were superior to all of the pictures in the other Battery.

Order of presentation again is not significant as a main effect.

The non-significance of the main effects, Batteries and Blocks, gives additional evidence of homogeneity of the data and stability of the experiment.

5. *Analysis of Variance for Length*

Some of the features of the length of the stories were presented in Section III. To determine whether length as a characteristic of the stories made the same differentiations in the data as the four scoring methods just discussed, an analysis of variance was made. The following hypotheses were rejected at the 1 per cent level.

Main effects:	Situations, Methods
Second order interactions:	Order x Method x Battery
First order interaction:	Order x Blocks
	Order x Method
	Blocks x Method
	Blocks x Battery

The outstanding feature of these results is the agreement between length and the four other scoring methods in the results pertaining to the main effects. The means among situations and between methods are significantly different and the means between blocks, orders of presentation, and batteries are not significant. Among the interactions some agree and others disagree with the interaction results of the other scoring techniques.

The methods means are significantly different in favor of the motion picture. These results suggest the possibility that the significantly greater amount of psychologically interpretable material produced by the motion picture as method and by some situations more than others, may be due to the fact that the stories were longer. There is evidence to support this probability in that the percentages of the total amount of psychological variables—need-press, intraceptive language and rating scale scores—to the total productivities for motion and still pictures are not significantly different. This may suggest that the stories were qualitatively similar. A more precise determination of the relationship between length and these variables, story by story, could be made by covariance techniques. The emphasis in this research has been which method produced quantitatively more psychologically significant material?

E. Synthesis of Results of Analyses of Variance and Other Evidence of Consistency in Analysis

The purpose of making four different analyses of the data was to obtain evidence of the extent of consistency in testing the various hypotheses. A summary of the tests of significance is given in Table 15. The outstanding feature of this table is the consistency in the results of the four methods in testing the main effects hypotheses.

TABLE 15
SUMMARY OF TESTS OF SIGNIFICANCE FOR FOUR METHODS OF ANALYSIS

	Need-Press	Intraceptive language	Discomfort relief	Rating scale
Main effects				
Order	—	—	—	—
Blocks	—	—	—	—
Situations	**	**	**	**
Method	**	**	—	**
Battery	—	—	—	—
1st order interactions				
Order x Blocks	**	—	—	**
Order x Method	**	—	*	—
Order x Battery	—	—	—	*
Blocks x Method	—	—	—	*
Blocks x Battery	**	—	*	—
Situations x Method	—	—	*	**
Situations x Battery	*	**	*	**
Method x Battery	**	—	—	**
2nd order interactions				
Order x Blocks x Battery	**	**	—	**
Order x Blocks x Situation	—	**	*	**
Order x Method x Situation	—	—	—	*
Blocks x Situations x Battery	—	—	—	*

*One asterisk (*) denotes significance at the 5 per cent level. Two asterisks (**) denote significance at the 1 per cent level.

The differences between the means for order of presentation, blocks, and batteries are not significant throughout. The differences among the situations means are significant by all of the analyses. The differences between the methods means are significant by all analyses except Discomfort-relief words. The fact previously demonstrated that the assumption of homogeneity of variance cannot be made with confidence with respect to this analysis may be sufficient grounds for giving less weight to its results.

The lack of consistency among the interactions for the different methods of scoring makes it necessary to be cautious in interpreting the significance of these measures. Since interactions are based upon a smaller number of

measures than main factors, decisions about their significance should be carefully made (13). In the previous discussions probable interpretations of the meaning of significant interactions were made. It should be emphasized that these interpretations are highly tentative. Kendall (21) sounds a word of warning about the meaning of interactions. "It must not be overlooked that significant interactions do not necessarily imply interaction in any real sense. They may arise from heterogeneity in the data."

How may the results of the main effects be interpreted? It has been suggested above that blocks, batteries, and order of presentation represent structural features of the experimental design through which we seek to answer the basic questions about situations and methods.

Blocks represent the replicating aspect of the design, that is, through it we get some evidence of the stability of the experiment. The fact that the blocks differences are consistently non-significant adds confidence to the inferences we may make about situations and methods.

The batteries differences are also consistently non-significant. This fact provides evidence of the homogeneity of the groupings of subjects in addition to the tests of homogeneity of variance previously applied in connection with testing the significance of the situations means.

The order of presentation differences are consistently non-significant. We may interpret this evidence to mean that the significant differences between methods and among situations are independent of order of presentation as defined in this research.

Let us consider situations by looking at the evidence of consistency among the methods of scoring in making differentiations among the picture situation means. The mean values for each situation were presented in Tables 8, 10, 12, and 14. We may ask the question: to what extent did a given mean tend to have the same relative position in the ranking of mean values for the various methods of scoring. Since the assumption of homogeneity of variance cannot reasonably be made with respect to the Discomfort-relief word method the following presentations will be concerned only with the Need-press, intraceptive language, and rating scale methods of scoring.

Two methods of examining evidence of consistency were tried. The first approach we may call ranking by the magnitude of the means and the second, ranking according to the statistically significant differences between means. The rank of a given mean for the second approach was determined in the following manner. Each mean was compared with each of the other 11 means. The number of instances in which that particular mean was significantly greater than the other mean was noted. For example, in Table 8,

the mean of Situation 1 was compared with the other 11 means and found to be significantly greater than nine of the means. Since no other mean was significantly in excess of other means as many as nine times, Situation 1 received the rank of one. This procedure was used to determine the rank for each mean.

It has been observed above that the need-press and intraceptive language analyses are analytical approaches whereas the rating scale method involved scoring a whole story in terms of its position on a continuum. One measure of consistency may be the extent of agreement between these approaches in ranking the picture situations in order of amount of interpretable material produced. The rankings of the situation means for need-press and intraceptive language were combined and the resulting ranking was compared with the rating scale ranking by the method of rank-difference correlation. Both methods of ranking were used. The results are shown in Table 16. All rho values are significant at the 1 per cent level. This evidence supports the conclusion that the extent of agreement between these two scoring approaches in ranking the situation means is significantly greater than zero.

TABLE 16

RHO VALUES FOR COMPARISONS OF COMBINED NEED-PRESS AND INTRACEPTIVE LANGUAGE RANKING WITH RATING SCALE RANKING

Method of ranking	Situations (12)[a]	Situations (24)[b]
Magnitude of means	.929	.901
Difference between means	.967	.882

[a]Situations (12) refers to the 12 picture situations, each in two forms, the motion and still picture.

[b]Situations (24) refers to the 12 picture situations, each motion and still picture being considered separately.

That there was little or no difference between the two ranking methods is shown in the following way. A combined ranking, including need-press, intraceptive language, and rating scale methods of scoring, was determined by the magnitude of means and the difference between means methods. These two sets of rankings were then correlated. The rho value for 12 situations is .997, for 24 situations .967. The final summary of the ranking of picture situations is shown in Tables 17 and 18.

The question might then be asked: how did the motion picture situations compare with their equivalent still pictures in power to evoke fantasy material? It can be seen from Table 18 that there is a tendency for a motion picture and its equivalent still picture to occupy similar positions within its own group. If we compare the rankings of the 12 motion picture situations

with their equivalent still pictures by the rank-difference method, we obtain a rho value of .839. This value represents the ranking of the combined magnitude of means and difference between means methods. Separately the former gave a rho value of .869 and the latter a rho value of .855. All of these values are significant at the 1 per cent level. These results indicate a tendency for a still picture to rank about the same as the motion picture depicting the same situation. Another possible interpretation may be that the situation depicted may be of greater importance than its depiction in motion or still picture form.

Another measure of consistency in differentiating among situations was obtained through a comparison of the rankings of the situations according to a measure of psychological significance and a ranking by a measure of non-psychological significance. It will be recalled that the rating scale analysis required the rating of the stories on a "description-projection" continuum. Stories rated one or two may be considered to be much more descriptive than those rated three or four. If we obtain the frequencies of one and two ratings for each situation and combine them, we obtain a score which can be taken as a measure of description, or a score of relatively non-psychological significance.

The situations were then ranked in descending order, the situation with the highest incidence of stories rated one and two receiving rank one, etc. This ranking of the situations was compared with a ranking based on psychological meaning (combined need-press and intraceptive language ranking) by the rank-difference correlation method. The obtained rho value was minus .05. This finding is interpreted to mean that a situation which stimulated many stories high in psychological significance produced few stories high in descriptive content.

We have presented evidence of considerable consistency among the methods of scoring in differentiating among the picture situations. In view of the final rankings of the situations in Table 17 we may interpret the significance of the Situations x Battery interactions shown in Table 15 to be attributable to the fact that both high ranking and low ranking situations were found in each battery. This evidence will be of value in discussion of the content characteristics of the stories and in making comparison between effective and non-effective pictures in Section V.

The meaning of the differences between the methods may be best understood within the context of the findings discussed so far. It has been shown that the picture situations tended to rank in somewhat the same order within the groups of motion and still pictures. In other words an effective motion

TABLE 17
FINAL RANK OF TWELVE PICTURE SITUATIONS

Situation	Rank
1	1
2	10
3	6
4	12
5	5
6	11
7	3
8	8
9	7
10	9
11	2
12	4

picture tended to be an effective still picture. The fact remains, however, that the motion picture did produce more of the psychological variables measured than the still picture.

The significance of the interaction of Method x Battery was previously indicated to be in large part attributable to the differences between methods found in Battery *B*. This may now be explained in terms of the evidence in Table 18. The motion pictures 7, 8, 9, 10, 11, and 12 were in Battery *B* and their combined rankings are less than those of the pictures in Battery *A*. In other words a greater number of the more effective pictures were in Battery *B* than in Battery *A*. Of course the confidence with which this inference can be made is dependent upon the assumption that the individual differences among the subjects were randomly distributed between the two batteries.

Even a casual inspection of the raw data revealed that all subjects did not respond in the same way to the two methods. It is likely that some subjects responded better to motion pictures, some to still pictures, and others about the same way to both. This possibility would provide another interesting focus for inquiry.

F. SUMMARY

In this section we have considered the methodological questions which constitute the core of this inquiry. Four methods of scoring were described. The data resulting from the application of these methods to the raw data were subjected to analysis of variance techniques. In addition an analysis of variance was run on length considered as a scoring method.

The results of these procedures were presented and discussed separately

TABLE 18
FINAL RANK OF TWENTY-FOUR PICTURE SITUATIONS

Situation	Rank
Motion Pictures	
1*a**	1
2*b*	15
3*c*	9
4*d*	21
5*e*	7
6*f*	20
7*g*	8
8*h*	11
9*i*	10
10*j*	13.5
11*k*	2
12*l*	4
Still Pictures	
13*g*	6
14*h*	18
15*i*	16
16*k*	5
17*j*	17
18*l*	19
19*a*	3
20*b*	24
21*c*	13.5
22*d*	23
23*e*	12
24*f*	22

*The same letter indicates same situation, e.g., 1*a* and 19*a* are same situation, one in motion, the other in still form.

for each scoring method. Finally a synthesis of the results of these various methods was attempted.

The evidence presented in this section supports five major conclusions. We should be reminded that the demonstrated tenability of such assumptions as randomisation of individual differences and the homogeneity of variance within the factorial design used adds to our confidence in the results. In addition the structural feature of blocks within the design provided replication of the experiment. The consistent non-significance of blocks differences increases the confidence which may be placed in the stability of the experiment.

On the other hand the instability of the significances of the interactions of various factors suggests caution in the interpretation of the main factors as "pure" effects. The major conclusions are:

1. The various methods used for scoring the records showed appreciable consistency in testing the major hypotheses.

2. The motion picture method mean was significantly greater than the still picture method mean rather consistently in power to produce psychologically meaningful material.

3. Significant differences between the picture situation means in stimulus power were consistently identified.

4. An effective motion picture tended to be an effective still picture. It may be that motion in a picture cannot be separated from the properties of the total picture situation or, motion is an integral part of the total situation.

5. The use of analysis of variance in factorial design has been demonstrated to be an appropriate technique in an exploratory study of projective methods.

V. CONTENT CHARACTERISTICS OF THE STORIES AND PICTURES

The presentation so far has been concerned with the methodological aspects of the research. The purpose of this section is to report evidence about the content characteristics of the stories and pictures. It will be recalled that the attempt was made to channel the subjects' fantasies into the area of family relationships by constructing pictures depicting such relationships and by verbal structuring when administering the tests. The results of this effort will be considered under the heading Effects of Verbal Structuring on Stories.

Another aspect of content will be revealed in answer to the question, what kinds of topics were introduced into the stories? This might be called manifest content. It is discussed as Analysis of Stories by Specific Content.

In contrast to the manifest or obvious content is latent content or those aspects of the stories which reveal the feelings, strivings, and attitude of the subjects. These aspects are made accessible to interpretation through some scheme of psychological analysis, in this instance, need-press. It is discussed under The Need-Press Distribution Among Stories.

Finally, we will consider some of the characteristics of the most and least effective of the pictures in terms of emotional tone, the activity and motion in the pictures, and the persons depicted in the pictures.

A. Effects of Verbal Structuring on Stories

The verbal instructions given to the subjects differed from the usual procedures in administering thematic apperception tests by specifically naming the general class of persons represented in the pictures. This procedure represented an attempt to channel the fantasy expressions into the area of family relationships. It is recognized that according to projective test theory the individual always behaves as a whole in any given situation. This means that the individual will not only reveal something about family relationships if that is the specific orientation of the story; he also reveals something about himself regardless of the specific content structure of the story.

Tomkins (39) claims that direct references to parental figures in *TAT* stories are not common. In making inferences about "the region of the family," to use Tomkins' words, the interpreter uses such clues as ambiguous figures interpreted as parental figures, older figures interpreted as parental figures, omission of older adults in the picture, number of stories about the family, length of such stories, and the intensity of their affect.

Consistent with the methodological emphasis of this study the data have been used to answer two questions. (*a*) How did the subjects respond to the verbal structuring—to what extent did they conform to the instructions by including in their stories the family figures? (*b*) What persons in addition to those given in the instructions were introduced into the stories?

TABLE 19
EXTENT OF CONFORMITY TO VERBAL INSTRUCTIONS BY SITUATIONS

Situation	*A*[a]	*B*[b]	*C*[c]	*D*[d]	Total
1	34	12	4	0	50
2	28	16	0	6	50
3	16	0	34	0	50
4	34	3	6	7	50
5	29	18	0	3	50
6	21	4	25	0	50
7	43	7	0	0	50
8	33	5	5	7	50
9	41	6	1	2	50
10	39	3	6	2	50
11	35	11	4	0	50
12	38	9	1	2	50
Total	391	94	86	29	600
Percentage	65	16	14	5	

[a]All persons given in instructions are mentioned in the stories.
[b]Some persons are mentioned, others are not.
[c]Used separate but appropriate pronouns instead of specific persons.
[d]Antecedents of pronouns are indefinite.

It is evident from Table 19 that the degree of conformity to the instructions was very high. In 65 per cent of all the stories the persons given in the instructions were specifically mentioned in the stories. In 81 per cent of the stories there was complete or partial adherence to the instructions. In 95 per cent of the stories it was clear that the stories were developed around the family characters given in the instructions. In only 5 per cent of the stories was it uncertain whether the story characters were the same as given in the instructions.

Additional evidence of the effects of verbal structuring is presented in terms of the frequency with which persons not given in the instructions were introduced into the stories. Out of the 155 persons who were introduced 96, or 61 per cent, may be classified as family characters.

This evidence indicates that insofar as the character composition of the stories is concerned the verbal structuring was highly successful. It also supports the possibility that fantasies may be directed into areas which may be the particular focus of investigation. It should be recognized that this

might have been achieved without direct verbal instructions about who is to be seen, since the situations were intentionally restricted by including only adult and adolescent figures. It is unlikely, however, that the high degree of conformity achieved here would have resulted.

B. Analysis of Stories by Specific Content

Some subjects gave stories without specific content structure while others told elaborate stories woven within the context of the school situation. The specific topics introduced into the stories were tabulated as follows. An empirical approach was followed in determining the scheme for classification of the topics. For example, if a person said, "the boy has done something," no tally was made but "the boy asked for money" was listed as "money." A particular topic was listed only once for a given story regardless of the frequency of mention within that story.

For summarizing the data in this analysis the tallies for sub-topics were combined to give a total for the more general category, as shown in Table 20. The data may be interpreted, for example, by noting that 84 times in 576 stories or in 15 per cent, school was a specific topic. It is evident, too, that the frequency with which school affairs were discussed was well distributed in the reactions to the various picture situations.

The topics of most frequent occurrence were school, finance, recreation and social, dates, personal habits, and car. It is interesting to compare the content of these stories with other reported studies in which the intent was

TABLE 20
Frequency of Topics Occurring in 576 Stories (By Situations)

Topics	Situations													
	1	2	3	4	5	6	7	8	9	10	11	12	Total	%
1. School	9	7	10	1	7	9	12	3	8	9	4	5	84	15
2. Financial	3	1	0	3	0	0	2	6	5	4	12	6	42	7
3. Recreation & Social	1	2	1	2	3	10	1	4	1	4	1	5	35	6
4. Dates	1	1	2	1	4	4	5	1	2	4	5	0	30	5
5. Personal habits	1	1	3	0	6	0	1	2	0	1	6	4	25	4
6. Car	1	1	0	1	1	0	8	2	0	0	7	3	24	4
7. Travel	0	5	1	0	2	1	1	2	0	0	1	1	14	2
8. Occupation	3	0	3	0	0	0	2	2	1	0	0	3	14	2
9. Law	3	2	0	0	0	0	2	0	0	0	1	2	10	2
10. Holidays	0	0	0	0	0	5	0	1	0	0	1	0	7	1
Total	22	20	20	8	23	29	34	23	17	22	38	29	285	

to determine sources of disagreement between adolescent boys and their parents (23). There is considerable commonality between these studies and the data presented in Table 20. These data tend to support the evidence presented previously that the stories were highly "family" oriented.

Caution should be exercised in making inferences about these data because it is evident from Table 20 that only about one-half of the stories had specific content. For example, a story built around such a theme as "the boy has done something wrong and is being scolded by his mother" would be quite meaningful but would not be listed as having specific topic content.

C. Need-Press Distribution Among Stories

The manifest content analysis reveals the specific topics which were most commonly found in the total group of stories. Another interesting kind of information may be obtained as a by-product of the need-press analysis This is the frequency of occurrence of the most common needs and press for each picture situation.[12] We may remember that needs describe forces within the personality, drives from within. The defined needs obviously cannot be interpreted as separate drives.

The drive toward ascendance or independence or defense of self is important as may be inferred from the high frequency of such needs as aggression (rank 4), autonomy (rank 5), acquisition (rank 8), and defendance (rank 10). Sanford (1) calls these, with a few others, the "ascendance" syndrome. The need succorance is found in the "dependence" syndrome, the need abasement in the "guilt or remorse" syndrome. This is just a way of pointing out that the configurations of needs found in stories may make sense within a particular conceptual framework. They may also be accessible to a more empirical interpretation. The data support the frequent statements about the independence-dependence struggle in adolescence. They may also be viewed as indicating one of the conflicts involved in achieving a developmental task. It is notable that the "self-sufficiency syndrome" including needs, seclusion, and rejection does not rank very high among the patterns found.

The distribution of press, which have been defined as outside forces impinging upon the individual, indicates that conflicting forces are high in frequency. For example, the controlling influences such as dominance, aggression, and rejection ranked 1, 2, and 3 in frequency, while accepting and helpful influences such as nurturance and exposition are also high. How-

[12]For detailed information about these data see the dissertation on file at the University of Chicago.

ever, the data suggest that considerable environmental pressures are reflected in the fantasies. It may well be that this is to be expected since projections frequently reflect forces that are unacceptable or unrecognized by the subject. This does not surprise us for the sample population comes from a middle class culture which exerts formidable parental pressures on adolescents. "Indeed, the culture of middle class Europeans and Americans probably exerts more severe pressure upon the young child—upon both his bodily processes and his emotional development—than does the culture of any other people in the world" (8).

It is apparent from the data that particular needs and press are elicited more frequently by some situations than by others. For example, Situation 3, which depicts a boy alone, elicited considerably fewer press than pictures where adults are present, such as Situations 1, 7, and 11. And as has been demonstrated, some situations tend to elicit more meaningful psychological material, in terms of combined needs and press of all kinds, because they involved the subjects more readily.

D. Effective and Non-Effective Pictures

The relative effectiveness and non-effectiveness of the various picture situations will be examined by considering three characteristics.

1. *Emotional Tone of the Pictures*

After the pictures had been developed, the investigator rated each one according to emotional tone. This was simply an effort to describe in a crude way another characteristic of the pictures. These ratings were given in Table 1.

It is apparent from Table 21 that the category marked "negative" contains the motion pictures with the highest rankings. These data support the recognized fact that most of the pictures in the *TAT* are of this negative quality. There must be enough emotional attractiveness in a picture

TABLE 21
Relationship between Emotional Tone of the Motion Picture Situations and Their Ranking by Effectiveness

Positive	Negative	Neutral
8	1	5
10	2	6
11	3	12
	4	
	7	
	9	

to involve the subject. Rapaport (34) conducted an item analysis of the *TAT* cards and describes for each card what he calls "the specific significance of the card." This refers to the kinds of feelings, and attitudes towards persons and objects which the various cards tended to elicit. One need only read them to be impressed with the fact that the data are "negative." This is consistent with the viewpoint in projective test theory that the more positive, acceptable feelings and attitudes are more readily expressed in everyday behavior and that projective tests purport to uncover tendencies which the individual dare not or cannot reveal.

2. *Activity and Motion in the Pictures*

Unfortunately there is no objective criterion of the amount of motion or the intensity of motion in the pictures which might be used to judge their effects upon the productiveness of various pictures. In none of the pictures could the motion be described as violent or eruptive. There are conversations, gestural movements in speaking, postural movements of the body, and entrances and/or exits of the characters.

TABLE 22
COMPARISON OF THE RANKING OF EACH MOTION PICTURE WITH ITS EQUIVALENT STILL PICTURE

Situation	Motion picture	Still picture
1	1	3
2	15	21
3	9	13.5
4	24	23
5	7	12
6	20	22
7	8	6
8	11	18
9	10	16
10	13.5	17
11	2	5
12	4	19

An attempt was made to relate the categories of description of the pictures used in Table 2 to their final productiveness. The notion here was that the complexity of the picture, as defined by the number of positional and attitudinal changes, might be an indication of the amount of motion in the picture. These categories of change ranged from four to 11. The pictures were ranked, the one with the least number of categories receiving rank one. These rankings were compared with the effectiveness ranking of the pictures. The obtained rho value was minus .48. This might suggest that the more

complex the picture, as here defined, the more effective the picture. However the obtained value was not reliable.

We have already observed that a good motion picture tends to be a good still picture. If we compare the rankings of the motion pictures with their equivalent still pictures we find from Table 22 that in 10 of the 12 situations, the motion picture had the higher rank. The two exceptions were Situations 4 and 7 where the rank differences were one and two respectively.

3. *Persons Depicted in the Pictures*

In order to present in sharp relief the situations which were productive and those that were much less productive, Table 23 has been prepared.

The widely accepted notion that pictures are most effective when there is someone in them with whom the subject can identify is supported by the data. The only pictures without an adolescent figure ranked 8 and 12 in effectiveness. This does not surprise us for this criterion for effective pictures was applied in the construction of the pictures.

TABLE 23
SUMMARY OF RANKINGS OF SITUATIONS ACCORDING TO EFFECTIVENESS AND NON-EFFECTIVENESS

Situation	Characters	"Effective" Rank[a]	Non-effective" Rank[b]
1	Woman-boy	1	12
2	Man-woman-boy-girl	10	3
3	Boy	6	7
4	Man-woman	12	1
5	Man-woman-boy	5	8
6	Boy-girl	11	2
7	Man-boy	3	10
8	Man-woman	8	5.5
9	Man-boy-girl	7	5.5
10	Boy-girl	9	4
11	Woman-boy	2	11
12	Man-woman-boy	4	9

[a]Ranking according to psychological significance.
[b]Ranking according to production of descriptive stories—high rank, e.g. (1) represents high production of descriptive stories.

Another fact of interest is that the three highest ranking pictures included one parent and an adolescent boy. The two pictures depicting a boy and girl ranked 9 and 11, while the two pictures including a man and woman ranked 8 and 12. There is little doubt that the pictures involving the adolescent *and* parental figures were the most effective.

E. Summary

The evidence in this section supports the following conclusions.

1. There is considerable evidence to support the belief that particular areas of personality such as family relationships may be studied by so structuring the stimulus situation as to produce data in that area.

2. The specific content of the stories reveals the rather typical concerns of adolescents of middle class culture as demonstrated by the comparability of the topics with other studies of adolescent-parent relationships.

3. The data about the psychological forces within the adolescent boy and the environmental pressures upon him as revealed in the need-press analysis are consistent with known facts.

4. The effective picture situation for eliciting fantasies was characterized by negative emotional tone, the presence of a boy and a parent, and a certain intensity of attitude which may be more or less explicit.

5. The least effective picture situation was characterized by positive or neutral emotional tone, the presence of a boy and a girl or parent, or by parents alone, and a lack of intensity in the implicit attitudes revealed in the situation.

VI. RELIABILITY OF THE STUDY

Walker (40) has defined reliability in the following terms. "If unchanging subjects are measured twice with a perfectly reliable instrument by a perfectly reliable agent, the correlation between the two sets of scores is 1.00." The chief sources of unreliability in a test are identified as the instrument, the subjects, and the investigator. It is recognized that unreliability or heterogeneity in the data of an experiment cannot easily be isolated as due to various ones or combinations of these elements. But it may be helpful for purposes of presentation to consider separately the issues related to these sources.

A. Reliability Related to the Instrument

Jackson (19) has pointed out that the idea of repeated measurement is important in the reliability concept. Since we have no test-retest condition in this research, it becomes necessary to get an estimate of reliability from the obtained data. A common way for obtaining such an estimate is the split-half method. This is based on an attempt to get a coefficient which gives the percentage of the obtained variance in a distribution of test scores that may be regarded as true variance, that is, as variance not due to the unreliability of the measuring instrument. Hoyt (18) developed a method using analysis of variance for obtaining an estimate of the discrepancy between the obtained variance and the true variance which was not subject to variations resulting from chance splittings of the test.

A coefficient of reliability is obtained by using the residual sum of squares as the basis for estimating the discrepancy between obtained variance and the true variance. The residual sum of squares is obtained by subtracting the "among subjects," "among stimuli," and any other significant factors or interactions from the total sum of squares. The residual sum of squares divided by the appropriate degrees of freedom gives the error variance. "This estimate of the discrepancy is a better one than that obtained by dividing the test into odd and even halves because in the latter case the particular split of the test, which is only one of many possible ways of splitting a test, may be an unlucky division and may result in either an overestimate or an underestimate of the coefficient of reliability" (18).

The coefficient of reliability may be obtained by the formula:

$$\frac{\text{"among subjects" variance-error variance}}{\text{"among subjects" variance}}$$

The coefficients of reliability calculated by this procedure for four scoring methods—need-press, intraceptive language, discomfort-relief words, and length—are presented in Table 24. The 16 coefficients ranged from .69 to .96 with a mean at .84 and the median at .85. The mean coefficient for each method was need-press .85, intraceptive language .79, discomfort-relief .82, and length .93. These values reflect a

TABLE 24
RELIABILITY COEFFICIENTS FOR GROUPS OBTAINED BY ANALYSIS OF VARIANCE

Group	Need-Press	Intraceptive language	Discomfort relief	Length
AMS	.89	.79	.81	.96
ASM	.80	.83	.79	.90
BMS	.88	.69	.78	.92
BSM	.83	.87	.88	.94

substantial degree of internal consistency among the groups by all methods of scoring. Referring to the work of Kuder-Richardson and Hoyt, Jackson contends: "The coefficients derived by the methods suggested by these writers are indices descriptive of the internal consistency of tests. The relationships between such coefficients and the reliability of tests obtained by test-retest methods requires to be established" (19).

B. RELIABILITY RELATED TO THE SUBJECTS

Since we have no data by retest of the same population over a time period, we will look at the probability of obtaining similar data on successive administrations of the test. We may look at past experience with thematic apperception material for evidence. According to Tomkins repeat reliability depends upon the basic stability of the personality. "The more stable the personality the more we should expect repetitions of the test to yield similar protocols" (39). This statement is qualified, however, by the following:

> Differences between successive administrations of the test may then be expected either if the individual is changing rapidly or if the change is slower but the test is repeated after a long period of time. Finally, we might well expect a decrease in repeat reliability if experimentally induced changes are introduced between successive administrations of the *TAT* (39).

Experimental evidence is presented to support these conclusions.

However, there is evidence that for developing adolescents the assumption of stability of personality cannot reasonably be made. According to Tomkins:

> Childhood and adolescence represent the period of maximal plasticity of the individual—the organism is never again so malleable. We should therefore expect successive administrations of the *T.A.T.* during this period to yield the lowest repeat reliability. This is in fact the case. Sanford reported an average repeat reliability of .46 of the needs expressed by children and adolescents who had been given the *T.A.T.* at yearly intervals over a three-year period (39).

Our best estimate of the probability of obtaining similar material on a repeat administration of these tests after a period of time is that there would be considerable change. If adolescents change we would expect their productions in a projective test to change. Another source of unreliability attributable to subjects is the mental set toward the test (32). The only evidence on this point in the present study is the impression of the investigator that the subjects generally related themselves readily to their tasks.

C. Reliability Related to the Scoring

In Section V the question of the reliability of each method used for scoring the records was discussed. It was then concluded that reasonable standards were achieved in terms of the consistency with which scores were given stories on repeated scoring. It should be recognized, however, that the measures indicated that the scoring procedures in addition to the instrument and the subjects constitute a source of unreliability in the results.

VII. CONCLUSIONS, IMPLICATIONS FOR PERSONALITY STUDY, AND SUGGESTIONS FOR FURTHER RESEARCH

A. Conclusions

The chief objective of this study was to explore the possibility of using the motion picture as a projective instrument for studying personality. The likelihood that an observer's perception of a motion picture might be sufficiently different from his perception of a still picture to produce results of interest in the field of personality study gave impetus to this research.

In order to explore this problem experimentally two important technical problems were solved. (*a*) Motion pictures which met important criteria for projective instruments were developed. (*b*) An experimental design for gathering and analyzing data in the most precise way was achieved.

The solution of these problems provided data in support of conclusions in which confidence can be placed because of the following characteristics of the research.

1. The assumptions underlying the techniques for making comparisons between groups were carefully observed. This applies particularly to the assumptions of homogeneity of variance and of the randomisation of individual differences.
2. The reliability indices of the instrument and of the scoring techniques compare favorably with standards in this field of inquiry.
3. A considerable degree of *consistency among methods* was achieved in testing the important hypotheses.

The conclusions warranted from the data are as follows:

1. The motion picture in general tends to elicit more material of psychological significance than the still picture. The generality of this conclusion is related to the assumptions underlying the scoring methods used and the concept of psychological significance adopted.

2. The motion picture can be used as an effective projective instrument for personality study.

3. The comparative stimulus values of different picture situations can be identified with an appreciable degree of consistency and accuracy.

4. There is a strong tendency for an effective motion picture to be an effective still picture. This finding suggests that a motion picture should be considered as a total stimulus situation rather than as a static picture to which motion has been added.

5. It is uncertain whether the superiority of the motion picture over the still picture method of presenting situations is of a magnitude to warrant the

development of a motion picture test for immediate use. Tests of significance only were applied to the data.

6. The use of analysis of variance in factorial design is demonstrated to be an appropriate technique in an exploratory study of projective methods.

7. The use of the wire recorder in obtaining data resulted in a more accurate record than otherwise obtainable and demonstrated that the testing time for this kind of test can be shortened considerably.

8. There is considerable evidence to support the belief that particular areas of personality functions such as intra-family relationships may be studied by so structuring the stimulus situation and test administration as to produce more data in that area than otherwise obtainable.

9. Many of the dominant aspects of the family culture of the particular group studied will be revealed in projective stories.

B. Implications for Personality Study

In addition to the conclusions based on evidence to test various hypotheses, the following implications for personality study are derived.

1. Review of the literature indicates that analysis of variance techniques in factorial design are here applied for the first time to projective test material. These procedures make possible the achievement of more rigorous standards of experimental technique than usually found in this field of inquiry and have considerable promise in studying questions related to: (*a*) Stimulus value differences among test items. (*b*) Study of group and individual differences. (*c*) Study of many variables on different levels, such as varying degrees of adjustment on several levels, including age, sex, and social status. (*d*) The identification and isolation of sources of heterogeneity which are sometimes ignored by research workers. For example, a "controlled" variable which is controlled by "arbitrary definition," to use Johnson's (20) term, will impair the quality of results.

2. The presence of investigator bias as a source of heterogeneity in data suggests the need for more replication in the research in this area of psychology. Attempts to repeat experiments may yield estimates of investigator error which cannot be derived from one study. This procedure would increase the quality of research products.

3. This study has implication for understanding the nature of audience participation and interest in commercial movies. It seems apparent that the projection of personal needs into the perception of motion pictures is an important clue to the reasons why movies are so attractive for adolescents.

4. It is evident that the important influences of family relationships in

the behavior of adolescents in school is becoming increasingly recognized by educators. Although effort was made to channel the stories into the area of family relationships, the most frequent content structure of stories was the school situation. This may reflect how inseparately the two are related.

5. The demonstration that the responses of the observer to the motion picture are projections of the student's personality has implication for the utilization of motion pictures as instructional media. The pictures which the writer used were intentionally ambiguous in order to induce a maximum amount of observer projection into his interpretations, but it is probably true that the observer always interprets a motion picture in terms of his own needs, strivings, goals, and aspirations. The instructor who is sensitive to this fact has a better chance of understanding student reactions to motion pictures in instructional situations.

C. Suggestions for Further Research

1. It is believed that the definition of motion as people talking and moving about has been given fairly adequate consideration by this study. Future research should extend the kinds of motion to include more dynamic interactions. A more careful control of motion, with some measure of the amount and kind seems indicated.

2. One type of motion has been used, namely, the movement of the persons in the picture situation. Another use of the motion picture would be to project the observer into the situation by use of the camera techniques suggested by Gibson (28).

3. The use of recording apparatus should be used when language analyses of the data are to be made. It was the impression from this research that the person recording by hand tends to impose his own language structure on the verbalizations of the subject.

4. The selective character of the tested population suggests an extension of this method to other kinds of groups. Longitudinal studies of different age groups in the normal adjustment range might be made. Another direction for research might include different clinical groups such as the maladjusted, the delinquent, or the extreme behavior deviate.

5. Individual differences among the subjects were not studied in this inquiry, although it is evident from the data that these differences are important data. This should provide a significant avenue for further study.

REFERENCES

1. Barker, R. G., Kounin, J. S., & Wright, H. F., *Eds.* Child Behavior and Development. New York: McGraw-Hill, 1943. Pp. 652.
2. Baxter, B. Problems of planning psychological experiments. *Amer. J. Psychol.*, 1941, **54**, 270-280.
3. ———. The problem of reliability in relation to factorial design. *J. Gen. Psychol.*, 1943, **29**, 157-161.
4. Bellak, L. An experimental investigation of projection. *Psychol. Bull.*, 1942, **39**, 489.
5. Boring, E. G. A History of Experimental Psychology. New York: Century, 1929. Pp. 699.
6. Crutchfield, R. S. Efficient factorial design and analysis of variance illustrated in psychological experimentation. *J. of Psychol.*, 1938, **5**, 339-346.
7. Dashiell, J. F. Fundamentals of General Psychology. Boston: Houghton Mifflin, 1937. Pp. 655.
8. Davis, W. A., & Havighurst, R. J. Father of the Man. Boston: Houghton Mifflin, 1947. Pp. 245.
9. Deabler, H. L. The use of Murray's need psychology in college personnel work. *Psychol. Bull.*, 1942, **39**, 431.
10. Despert, J. L., & Potter, H. W. Technical approaches used in the study and treatment of emotional problems in children. *Psychiat. Quart.*, 1936, **10**, 619-638.
11. Dollard, J., & Mowrer, O. H. A method of measuring tension in written documents. *J. Abn. & Soc. Psychol.*, 1947, **42**, 3-32.
12. Dysinger, W. S., & Ruckmick, C. A. The Emotional Responses of Children to the Motion Picture Situation. New York: Macmillan, 1933. Pp. 122.
13. Fisher, R. A. The Design of Experiments. Edinburgh: Oliver & Boyd, 1937. Pp. 260.
14. ———. Statistical Methods for Research Workers. (Seventh ed.) Edinburgh: Oliver & Boyd, 1938. Pp. 356.
15. Frank, L. K. Projective methods for the study of personality. *J. of Psychol.*, 1939, **8**, 389-413.
16. Garrett, H. E. Statistics in Psychology and Education. (Third ed). New York: Longmans, Green, 1947. Pp. 487.
17. Henry, W. E. The thematic apperception technique in the study of culture-personality relations. *Genet. Psychol. Monog.*, 1947, **35**, 3-135.
18. Hoyt, C. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153-160.
19. Jackson, R. W. B., & Ferguson, G. A. Studies on the Reliability of Tests. Toronto: Univ. Toronto Press, 1941. Pp. 132.
20. Johnson, P. O., & Tsao, F. Factorial design in the determination of differential limen values. *Psychometrika*, 1944, **9**, 107-144.
21. Kendall, M. G. The Advanced Theory of Statistics. (Vol. 2) London: Griffin & Co., 1946. Pp. 572.
22. Köhler, W. Gestalt Psychology. New York: Liveright, 1929. Pp. 403.
23. Landis, P. H. Adolescence and Youth New York: McGraw-Hill, 1945. Pp. 470
24. Lindquist, E. F. Statistical Analysis in Educational Research. Boston: Houghton Mifflin, 1940. Pp. 266.
25. Masserman, J. H., & Balken, E. R. The psychoanalytic and psychiatric significance of phantasy. *Psychoanal. Rev.*, 1939, **26**, 515-519.
26. Monroe, W. S., Devoss, J. C., & Kelly, F. J. Educational Tests and Measurements. Boston: Houghton Mifflin, 1924. Pp 521.

27. MORGAN, C. D., & MURRAY, H. A. A method for investigating fantasies. *Arch. Neurol. & Psychiat.*, 1935, **34**, 289-306.

28. MOTION PICTURE TESTING AND RESEARCH. Army Air Forces Aviation Psychology Research Report No. 7. Washington: Gov't Printing Office, 1947. Pp. 267.

29. MURRAY, H. A. Explorations in Personality. New York: Oxford Univ. Press, 1938. Pp. 761.

30. ———. Techniques for a systematic investigation of fantasy. *J. of Psychol.*, 1936, **3**, 115-143.

31. ———. Thematic Apperception Test (and Manual). Cambridge: Harvard Univ. Press, 1943.

32. MURSELL, J. L. Psychological Testing. New York: Longmans, Green, 1947. Pp. 449.

33. PETERS, C. C., & VAN VORHEES, W. R. Statistical Procedures and Their Mathematical Bases. New York: McGraw-Hill, 1940. Pp. 516.

34. RAPAPORT, D. Diagnostic Psychological Testing. (Vol. 2.) Chicago: Yearbook, 1946. Pp. 516.

35. SANFORD, R. N., *et al.* Physique, personality and scholarship. *Monog. Soc. Res. Child Devel.*, 1943, **8**, No. 1.

36. SARASON, S. B. The use of the thematic apperception test with mentally deficient children. *Amer. J. Ment. Def.*, 1943, **47**, 414-421.

37. SNEDECOR, G. W. Calculation and Interpretation of Analysis of Variance and Covariance. Ames: Collegiate Press, 1934. Pp. 96.

38. SYMONDS, P. M. Criteria for the selection of pictures for the investigation of adolescent fantasies. *J. Abn. & Soc. Psychol.*, 1939, **34**, 271-274.

39. TOMKINS, S. S. The Thematic Apperception Test. New York: Grune & Stratton, 1947. Pp. 297.

40. WALKER, H. M. Elementary Statistical Methods. New York: Holt, 1943. Pp. 368.

41. WARREN, H. C., & CARMICHAEL, L. Elements of Human Psychology. Boston: Houghton Mifflin, 1930. Pp. 462.

42. WELLS, H. Differences between delinquent and non-delinquent boys as indicated by the *TAT*. *Psychol. Bull.*, 1945, **42**, 534.

43. WOODWORTH, R. S. Contemporary Schools of Psychology. New York: Ronald, 1931. Pp. 232.

44. WYATT, F. The scoring and analysis of the thematic apperception test. *J. of Psychol.*, 1947, **24**, 319-330.

Published as a separate and in *Genetic Psychology Monographs*, 1949, 39, 279-325.

THE ORGANIZATION OF HEREDITARY MAZE-BRIGHTNESS AND MAZE-DULLNESS* †

Department of Psychology, University of California, Berkeley

LLOYD V. SEARLE

	Acknowledgments	281
I.	Introduction	283
II.	Descriptions of apparatus	291
III.	Statistical analysis	301
IV.	Psychological analysis	311
V.	General summary	323
	References	325

*Received in the Editorial Office on January 15, 1949.

†A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Graduate Division of the University of California, Berkeley, in September, 1945.

ACKNOWLEDGMENT

The writer is indebted to Professor Edward C. Tolman for many suggestions and criticisms, and in particular, for his stimulating interest throughout the course of this research. Professor Warner Brown made a number of helpful suggestions for writing the original dissertation from which the present paper is drawn. Assistance and instruction were provided by Professor Clarence W. Brown during early stages of the experiment.

It is obvious that primary credit is due to Professor Robert C. Tryon, whose care in planning and conducting his own famous experiment is responsible for marked differences which occur in the performances of Brights and Dulls. The writer is indebted to him for suggesting the initial plan of this study, as well as for generously donating its subjects.

LLOYD V. SEARLE

111 Thirty-Fifth Street, S.E.
Washington 19, D. C.

I. INTRODUCTION

A. Purpose

The rats of Tryon's genetic strains have been selectively bred on the basis of their error scores in a complex alley maze, under conditions of rigorous environmental control, until two groups have resulted whose hereditary abilities to learn the maze are almost completely separate (18).

The Tryon study represents a significant experimental approach to the problem of heredity and environment. The trait of maze-learning ability has been generally accepted as the best available index of rat "intelligence," even though its psychological nature remains a matter of wide disagreement. In the results of this experiment it has been clearly shown that such a complex trait has a controllable hereditary basis. Specifically, the experiment succeeded in separating the genetic determiners into two sets, one favorable to maze-brightness and the other to maze-dullness. The gradual nature of separation of the strains through succeeding generations provides evidence that the independent factors involved are numerous (16).

These strains of rats constitute invaluable material for further investigation. Although it is pointless to argue whether heredity is more or less important than environment in the learning abilities of individual rats, it is clear that the existing difference between mean performances of these Bright and Dull groups is due wholly to genetic causes. Since maze error scores were used as the sole criterion of selection, this difference may be regarded as relating directly to the maze-learning function. Thus, any systematic differences occurring in various other behaviors of Brights and Dulls reflect the operation of the same sets of causes, and the common elements of such behaviors may be regarded as constituting basic psychological components of the maze-learning ability.

The present study undertakes to provide an analytical description of the total organization of behavior for each of the two strains, and in terms of the analyzed components, to present a preliminary explanation of the rats' maze-learning performances. Before the study was begun, evidence had already been obtained that the basic psychological difference was not to be conceived as a simple matter of ability to learn. In testing small samples of the Brights and Dulls in an exploratory series of problems, the writer had found that the Brights' superiority tended to disappear when stimulus conditions were made sufficiently different from those of the alley maze. In one case, involving a three-dimensional under-water maze, a slight but reliable difference was obtained in favor of the Dulls. Although results of these tests were

not conclusive, they were sufficient to indicate in a preliminary way that differences in a general factor of learning were either very slight or nonexistent. It might have been supposed, in line with conventional factor theory, that the results were explainable in terms of group or specific factors of learning, but this explanation did not seem tenable, either, in view of the striking consistency which both groups showed in certain of their "personality" patterns. It was observed that individual rats reacted in characteristic ways throughout widely varying problem situations. Although the responses were not systematically measured during the preliminary experiments, there were definite indications that reliable differences existed between Brights and Dulls in their emotional and motivational natures.[1] On the whole, the suggestion was strong that the groups represented distinctive personality *types*, whose contrasting degrees of maze-brightness were to be regarded as but one aspect of their different total organizations of behavior.

B. Statistical Design

In previous studies of trait organization in animals, the method generally employed has been that of intercorrelating scores by tests, where a variety of test measures has been obtained for a heterogeneous sample of the animals under study (2, 7, 13, 15, 22). Normally the data are treated by some method of factor analysis, and in recent years the concept of "trait organization" has come to be identified with this kind of experimental approach. For the present study, however, where the question of principal interest relates to differences in organization between two pure strains, the conventional approach could not be taken. The proposal that a separate analysis might be made for each strain was not seriously considered on the grounds that intercorrelations would tend generally towards zero to the extent that the strains were actually homogeneous, and also because differences which might exist between the two strains on various individual test measures would be lost in the process. More importantly, the subject of investigation is regarded as basically contradictory to the conventional assumption that trait organization is identical for all members of a given species.[2]

Based upon the above considerations, a decision was made to design the

[1] Emotionality studies of Brights and Dulls had been made previously by Tryon, Tryon, and Kuznets (20, 21).

[2] That is, that all individuals' performances in a given test involve the same set of component traits and in the same relative proportions. In multiple-factor methods this assumption is expressed by the formula, $x_a = a_1x_1 + a_2x_2 + \ldots + a_nx_n$, wherein the weights, $a_1, a_2, \ldots a_n$, may vary for other tests but in test a are constant for all individuals.

present experiment along the lines of techniques which have been employed by Stephenson (10) and by Zubin (24) for studying psychological "types" among humans. Briefly described, these techniques involve computing measures of relation among the various individuals, instead of among tests. Where a series of tests is considered to be a sample of behaviors in terms of which individuals show distinctive patterns, degrees of resemblance among individuals may be determined by intercorrelating the patterns, and as Stephenson has demonstrated (11), the intercorrelations may be treated by ordinary methods of factor analysis. The resulting factors—i.e., clusterings of variables—represent *types* of individuals.

Applied to an experiment with samples of Brights and Dulls, the method may be outlined in advance as involving the following steps: (*a*) A distribution of (comparable) scores is obtained for each rat in a series of test measures, where the series represents a sample of behaviors selected from the universe of rat behavior problems. (*b*) The correlation is computed between each pair of rats. For the total number, k, of Brights plus Dulls, the number of intercorrelations computed will be $k(k - 1)/2$. (*c*) Using some method of group factor analysis, the table of intercorrelations is analyzed for factors. Following these steps, the existence of hereditary behavior types might be considered demonstrated if factor saturations for individual members of the two strains were to provide evidence for a "Bright" factor and a "Dull" factor. A further identification of each type should then be made in terms of its characteristic pattern of behaviors.

In the course of designing the experiment, a problem was encountered initially in choosing a way to make the various test scores comparable. In raw form the scores would be obtained as numbers of errors, seconds, cage revolutions, etc., and would evidently have to be converted into units of some common scale before meaning could be attached to the distributions for individual rats. The previously suggested solutions to this problem have been either (*a*) to use only tests for which the units are already comparable in raw form, or (*b*) to convert the original scores in each test into units based upon the means and standard deviations of the experimental sample. The first solution is, of course, not applicable in the present case. The second was considered undesirable in view of the probable variation to be found among the forms of test distributions (many of these would be platykurtic or bimodal as a result of combining scores for Brights and Dulls), and also because intercorrelations among individuals are to some extent mathematically predetermined when test means and sigmas have been standardized with

reference to the experimental sample.[3] The decision was therefore made to employ a "reference" group of rats, representing the original population from which Brights and Dulls were bred, and to convert scores for rats of the experimental samples into a common scale based upon the normalized distributions of test scores for the reference group. Although complicating the experimental procedure in some respects, this method is free from the above objections and in addition, permits Brights' and Dulls' performances to be evaluated as "high" "low," or "average" on each test measure in terms which have a relatively standard meaning. It may be noted in passing that the principles of the method are identical to those which are involved in making up educational achievement profiles for school children, when the common scale is based on established "norms" for a set of tests (3).

C. Animals

The samples of Brights and Dulls were taken from the 22nd generation of the Tryon strains. They were members of a colony stock which had been removed from the regular experimental groups since the F_{10} generation, and for this reason had not been selectively bred on the basis of maze error scores since that time. They had been selected on the basis of coat color in order to prevent accidental intermixture of the two strains. The Brights had gray coats and the Dulls black.

For the reference group, a sample was chosen from a stock of rats known as "*M* by *M*" (*MxM*), so called because they were the progeny (F_7) of a median strain which had been introduced during an earlier generation by crossing members of the two pure strains. Presumably, rats bred in this way would resemble most closely the original parent population.

The three groups included a total of 40 rats. Of these, 12 were Brights, 12 were Dulls, and 10 were *MxM*'s. They were chosen from 10 litters of Brights, 10 of Dulls, and 16 of *MxM*'s (the total stock of the current generation), one rat being taken from each litter in order to provide the most representative samples obtainable with a small number of cases. Two Brights, two Dulls, and one *MxM* were initially designated to serve as spare replacements in the event that members of the experimental samples proper should be lost through sickness or accident. Only one of these five spares, G-6, was eventually included as a regular member.

The individual rats' numbers, coat color, ages, and weights are listed in Table 1. All rats were male and had pigmented eyes. As may be seen in

[3] The average intercorrelation cannot be less than the value, $-1/(k-1)$ (cf. ref. 1, p. 275), for a sample of k members, nor greater than zero.

TABLE 1

EXPERIMENTAL ANIMALS: NUMBER DESIGNATIONS, AGES, WEIGHTS, AND COAT COLOR
(*All rats were males with pigmented eyes*)

	No.	Coat Color	Age Begin (days)	Age End (days)	Wgt. Begin (grams)	Wt. End (grams)
Brights (F_{22})	G- 6	Gray	96	346	285	404
	G- 5	Gray	96	346	275	402
	G-10	Gray	99	349	270	420
	G-12	Gray	99	349	250	385
	G-16	Gray	98	348	290	404
	G-18	Gray	96	346	285	406
	G-24	Gray	91	341	310	414
	G-27	Gray	88	338	215	344
	G-28	Gray	76	326	275	434
	G-30	Gray	72	322	190	360
	Mean		91	341	264	401
Dulls (F_{22})	B- 3	Black	98	348	245	380
	B- 6	Black	96	346	230	316
	B- 8	Black	95	345	195	256
	B-10	Black	95	345	250	352
	B-12	Black	95	345	195	292
	B-15	Black	95	345	220	280
	B-18	Black	95	345	230	282
	B-22	Black	95	345	180	334
	B-24	Black	95	345	240	374
	B-26	Black	91	341	210	322
	Mean		95	345	220	319
MxMPs (F_7)	BH- 5	B.Hood	96	346	260	380
	BH- 9	B.Hood	82	332	240	356
	BH-11	B.Hood	96	346	270	374
	BH-13	B.Hood	99	349	210	370
	BH-17	B.Hood	96	346	260	360
	BH-22	B.Hood	93	343	265	378
	BH-26	B.Hood	101	351	225	282
	BH-30	B.Hood	99	349	250	376
	BH-33	B.Hood	101	351	210	340
	BH-37	B.Hood	96	346	235	344
	BH-41	B.Hood	99	349	270	310
	BH-45	B.Hood	95	345	210	374
	BH-48	B.Hood	84	334	210	364
	BH-54	B.Hood	84	334	195	360
	BH-56	B.Hood	72	322	255	434
	Mean		93	343	240	361

the tabulated figures, the groups were essentially equal in age, but their weights show a distinct positive correlation with hereditary maze-brightness.[4] With the exception of one Bright rat who died of a lung disorder, all rats maintained excellent health throughout the experiment.

[4] An unpublished study by Tryon indicates that a weight difference in favor of the Brights is characteristic.

D. General Procedure

During a period of eight months, 30 measures of the rats' behavior were obtained in nine different apparatus situations. These are listed in Table 2 in the order, from beginning to end of the experiment, in which each apparatus was employed.

TABLE 2
List of Variables

OPEN FIELD	1. Defecations
	2. Urinations
	3. Rearings
	4. Washings
	5. Distance
TRYON MAZE	6. Errors (2-19)
	7. Time (2-19)
	8. Errors (20-21)
	9. Time (20-21)
WATER TANK	10. Errors
	11. Time
	12. Starting Time
WATER UNIT	13. Starting Time
ELEVATED PATH	14. Time
14-UNIT MAZE	15. Errors
	16. Time
	17. Starting Time
REVOLVING CAGES	18. Total Activity
	19. Pre-feeding Activity
16-UNIT MAZE	20. Errors (2-4)
	21. Errors (13-15)
	22. Time (1-2)
	23. Time (6-12)
	24. Time (14-17)
	25. Starting Time
	26. *VTE*
6-UNIT DISCRIMINATION	27. Errors
	28. *VTE*
	29. Starting Time
	30. Time

Since the experiment had as its general aim the study of psychological differences resulting from genetic causes, it was considered of primary importance that no systematic environmental factors should be allowed to affect the strains differentially. Feeding, handling, living conditions, and exposure to the apparatus were therefore made as nearly equal as possible for all three groups. All of the rats were kept in a single rack of living cages, among which Brights, Dulls, and *MxM*'s were distributed systematically in order

to provide comparable physical and social living environments. Since single trials on some of the apparatuses required appreciable periods of time, it was impossible to make the presentation of problems exactly equal for all individual rats. It was also found impractical to randomize the order of running on each trial and on each piece of apparatus. Accordingly, each rat was assigned a rank at the beginning of the experiment and was run in that order, or its reverse, throughout. Constant differences among individual rats were introduced by this procedure, but since members of the strains were distributed, it served to maintain equal treatment of the groups and to facilitate the accurate recording of scores.

II. DESCRIPTIONS OF APPARATUS

A. Open Field

The Open Field apparatus for observing rats' emotionality responses was first described by Hall and Ballachey (5). As employed in the present experiment, it consisted of an 8-foot diameter circular area on the floor of the experimental room, enclosed by a sheet metal wall 12 inches high. The floor was marked off in numbered squares in order to facilitate making tracings on record sheets of the rats' paths within the area. Once a day for three consecutive days each rat was removed from his cage, placed in the center of the field, and allowed to remain there for a period of two minutes, during which time the measures listed below were taken:

1. *Defecations:* The total number of single faeces dropped.
2. *Urinations:* The number of separate urinations.
3. *Rearings:* The number of times the animal reared up on his hind legs.
4. *Washings:* The number of times the animal lifted forepaws to manipulate his face and whiskers.
5. *Distance:* Total distance represented by tracings of the rat's path, measured by chartometer.

B. Tryon Maze

The details of apparatus construction and procedures of running for the Tryon Maze have been described in a paper by Tolman, Tryon, and Jeffress (14). The standard procedure involves an initial 8-day period of preliminary training over a shortcut pathway (shown as a dashed line in Figure 3), followed by 19 days of running on the maze proper at one trial per day. The rats were given no food nor water throughout the entire period other than the reward which immediately followed each day's run, a soupy mixture of water and dry food. The present groups were run along with regular rats of the selective breeding experiment.

As an exploratory variation of the feeding procedure, two trials were added at the end of the 19th trial in which the rats were fed just before, instead of just after, the maze run. The automatic recordings of time and errors taken over all of the trials, 2 to 21, furnished data for the following variables:

6. *Errors:* Total number of full entrances into blind alleys during Trials 2 to 19, inclusive.
7. *Time:* Total number of time units, Trials 2 to 19.
8. *Errors:* Total number of blind alley entrances for Trials 20 and 21.
9. *Time:* Total number of time units, Trials 20 and 21.

C. Water Tank

The Water Tank apparatus was an 8-foot long galvanized iron tank, filled with water to a depth of 10 inches. The dimensions and essential parts are illustrated in Figure 1. The rats were placed by hand into the Starting Compartment, from which point they had to dive under the Starting Block and swim the length of the tank under water in order to escape. They were prevented from coming to the surface by a wire screen which extended nearly to the end of the tank, and were stimulated to remain close to the bottom by the tin "blocks" which hung from the screen at spaced intervals. The regions between these blocks constituted third-dimensional *culs-de-sac* which the rats learned gradually to eliminate over a long series of trials.

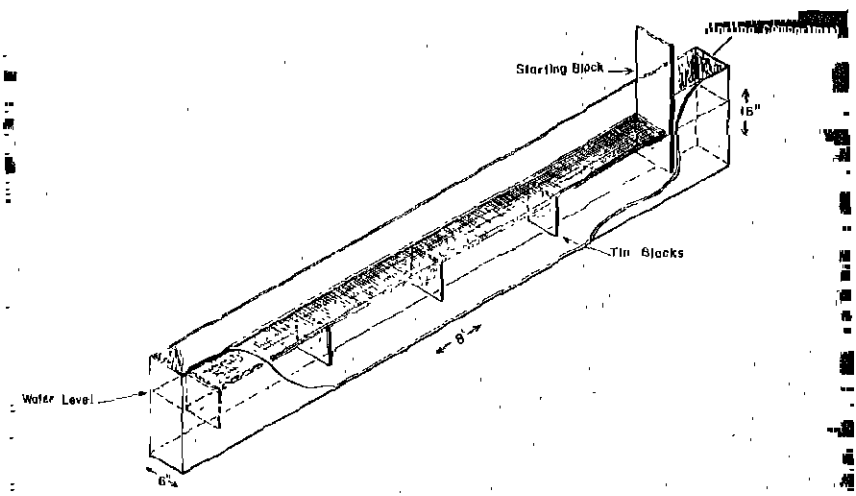


FIGURE 1
Water Tank

The procedure of running included a preliminary period during which the rats were trained gradually to swim under water. On the first trial they were placed into the water at the starting end and allowed to swim freely on the surface to a wire ladder leading out of the opposite end. Internal parts of the apparatus as shown in Figure 3 were suspended above the tank. Eight trials were given, during which the screen and blocks were lowered by successive stages into the water until, on the eighth trial, the screen was in its final position just under the surface. Eighteen further trials were then run, during which the following measures were obtained:

10. *Errors:* Each rat's score was taken as the sum of numerical values assigned to his entrances into the *culs-de-sac*. Full entrance up to the screen counted as "3," partial entrance as "2" or "1" depending on degree.

11. *Time:* Time was taken by stopwatch from the moment of diving under the surface to the moment of rising at the end.

12. *Starting Time:* After being placed into the Starting Compartment, the rats thrashed about or simply floated for some time before diving under water. These times were also recorded in seconds by stopwatch.

D. Water Unit

The Water Unit was a galvanized iron tank, similar in depth and width to the Water Tank described above but shaped in the form of a single *T*-maze unit. It contained water to a depth of 10 inches, and was so constructed that the rats could be made either to wade through or to swim through under water. Wading trials were made possible by a wire screen, placed four inches below the surface and extending to within six inches of the ends of stem, blind alley, and true path, with the remaining distance covered by tin gates (see Figure 2). For under-water trials these gates were raised on the stem and true path ends, opening the pathway under water and closing it off above. All ends of the tank were open above the 10-inch level, allowing the rats to enter and leave the apparatus just at the water surface. A starting compartment was attached at the stem end of the tank. This was a pan filled with water and covered over the top by a wire cloth cage, having a door in the rear for inserting the rat.

This apparatus was included as a problem in the present series because of the findings of a preliminary experiment in which Brights and Dulls had shown a marked difference in their readiness to dive voluntarily into the water. The procedure, which followed a preliminary training period on the Water Tank, involved giving the animals alternate wading and swimming trials in an insoluble right-left sequence, except that on each swimming (under water) run, the correct side was the same as on the immediately preceding wading run. Thus, a constant cue was provided which might have been learned, although none of the rats showed better than chance performance in this respect during the limited series of trials. The technique was to place each rat into the starting compartment by hand and allow him 60 seconds to enter the stem of the unit. If he had not entered at the end of that time he was given a light, intermittent sprinkling of water from above the compartment. Most rats would then dive off within a very few seconds. It was observed early in the experiment that on the first trial

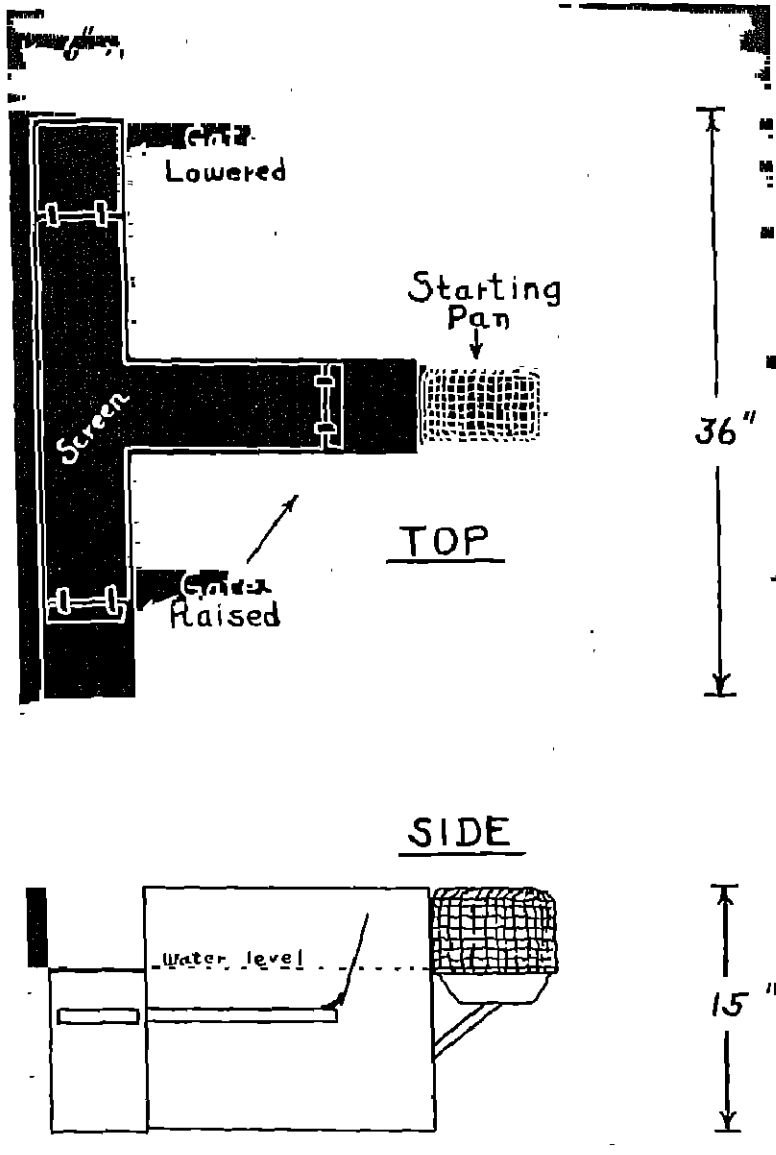


FIGURE 2
WATER UNIT

following their initial experience in making an under-water choice all members of the Dull group, without exception, left the starting compartment voluntarily, whereas the Brights, also without exception, remained the full 60 seconds and required the shower.

Essentially the same procedure was followed in the present experiment, and nearly identical results were obtained. Five wading and five under-water trials were given in alternate order.

13. *Starting Time:* The score of each rat was taken as the sum of starting times on under-water Trials 2 to 5, inclusive.

E. Elevated Path

The Elevated Path was 40 feet in length, two inches in width, and was mounted on standards to a height of 30 inches above the floor. It extended between two racks of sliding cages, one for starting the rats automatically and the other containing food as a reward at the end. A one-way gate was mounted on the path near the start and another near the end to prevent retracing.

During a series of preliminary trials the rats were accustomed gradually to the narrow pathway, gates, and end boxes. Following this period they were given eight trials at two trials per day. At the beginning of each trial they were placed in the sliding cage compartments at the start and were allowed to come out in turn to run the length of the path for food at the end. Times to run 40 feet between non-retrace gates were recorded by stopwatch.

14. *Time:* Each rat's score was the sum of his times in seconds for the eight trials.

F. Fourteen-Unit Maze

The pattern of the 14-Unit Elevated Maze is shown in Figure 3. Its pathway was two inches wide and 30 inches above the floor. An end-box rack, which is also shown in the diagram of Figure 3, had runners at two different levels for the end-boxes to slide on. The rats were first placed in individual compartments at the lower level, which connected to the pathway leading to the "Start" position. When the box was slid sideways for a distance of one compartment, a rat was freed to enter the maze. From the "End" position of the maze he entered a food compartment at the upper level. A total of 11 trials were given, during which the following measures were obtained:

15. *Errors:* Each rat's score was double the number of full and half errors during Trials 2 to 11, inclusive.

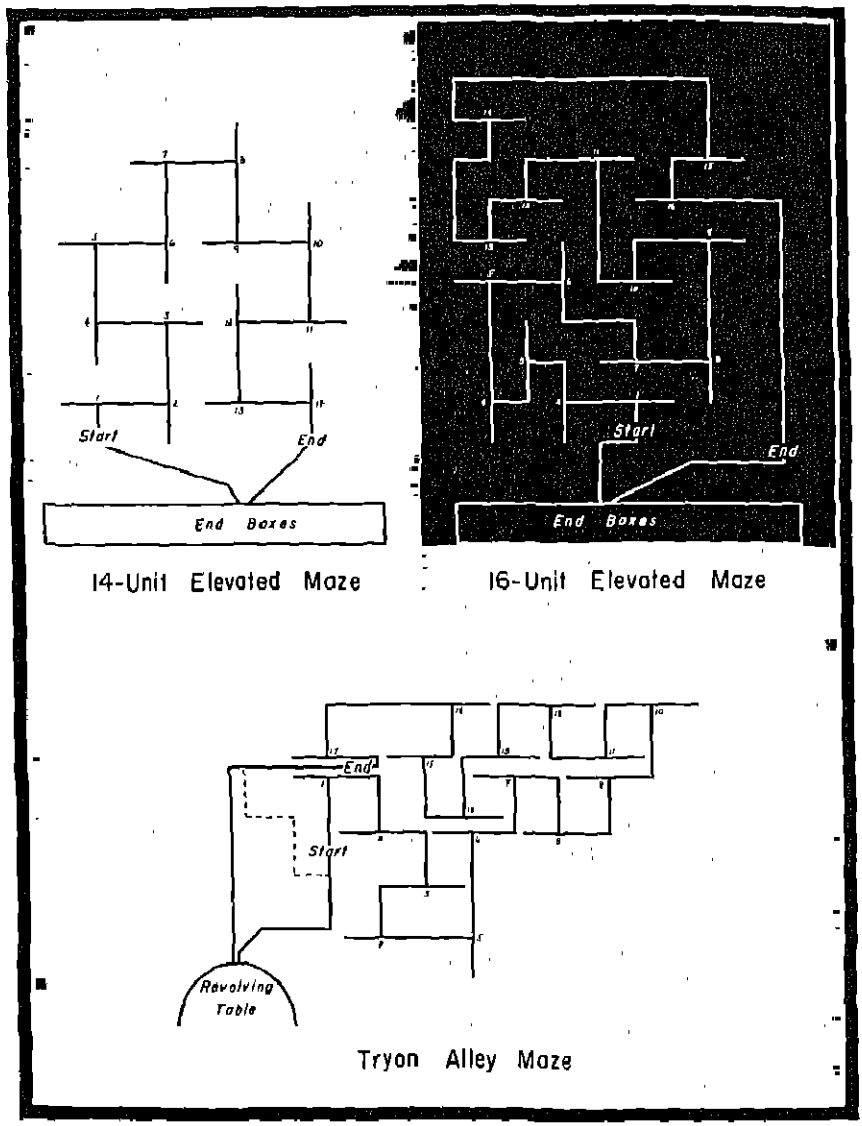


FIGURE 3
MAZE PATTERNS

16. *Time:* The total time in seconds from "Start" to "End," summed for the last 10 trials.

17. *Starting Time:* The time in seconds to come out of the starting compartment to "Start" position, summed for the last 10 trials.

G. Revolving Cages

The Revolving Cage apparatus was similar in general respects to others which have been commonly used for measuring rat activity levels. The wheels were mounted three tiers high on racks which stood against the walls of the experimental room. No stationary units were attached to the wheels. The details of their construction have been described in an earlier report (8).

The rats were not handled during the 15 days of running in the activity wheels. They were fed daily from 5:00 to 6:00 P.M., during which time the wheels were locked in a stationary position while their food dishes, containing wet mash, were left inside. The first eight days were allowed for "breaking-in," and thereafter counter readings were taken twice daily, at 3:00 and 5:00 P.M., for a period of seven days.

18. *Total Activity:* This measure was taken as the total number of revolutions for each rat during the seven days.

19. *Pre-Feeding Activity:* Scores were calculated for each rat as the ratio of activity during the two hours before feeding (i.e., 3:00 to 5:00 P.M.) to activity during the remaining 21-hour period of each day.

H. Sixteen-Unit Maze

The 16-Unit Elevated Maze was similar in construction to the 14-Unit Maze described above, but was of a more complex pattern. In Figure 3, patterns for both of these mazes and for the Tryon alley maze are shown for comparison. As in the case of the previous elevated maze, the rats started from boxes at the lower level of the end box stand and returned to food boxes at the upper level. Similar measures were taken of times and errors.

A total of 17 trials was given, at one trial per day. Error and time curves for all rats were found to plateau at about the 6th trial and remained relatively flat up to the 13th trial. At this point a variation in the room conditions was introduced in order to test the importance of visual cues in the rats' plateau performances. It had been noticed that the Dulls' error scores up to that time were very low, contrasting markedly with their Tryon Maze scores. Since visual factors are freer to operate in the elevated maze situation, this finding might have been expected on the basis of Krechevsky's theory that one of the fundamental differences between Brights and Dulls is a matter of

spatial and visual "hypotheses." If it should be found that Dulls were more disrupted than Brights by the *removal* of visual cues, additional support would be lent to Krechevsky's theory. Accordingly, it was decided to run a series of extra trials with the experimental room blacked out.[5]

Following the 12th trial, the room was completely sealed against external light and the rats were run for the remaining trials under two conditions of darkness: (*a*) On Trials 13 to 16, inclusive, a spotlight equipped with a red filter was used by the experimenter in order to follow the rats' paths around the maze and record errors.[6] (*b*) On Trial 17, an automatic switch was employed for recording total times and the rats were run in complete darkness. Both errors and times were recorded on Trials 13 to 15, and time only on Trials 16 and 17.

On the basis of learning curves for time and errors, the 17 trials were divided into initial, plateau, and test trials, and scores were summated separately for these three periods. The following measures were used:

20. *Errors:* Double the number of full plus half errors during Trials 2 to 4, inclusive.
21. *Errors:* Errors during Trials 13 to 15, inclusive.
22. *Time:* Total of times for Trials 1 and 2.
23. *Time:* Total of times for Trials 6 to 12, inclusive.
24. *Time:* Total of times for Trials 14 to 17, inclusive.
25. *Starting Time:* Total of starting times for Trials 14 and 15.
26. *VTE:* "Vicarious trials and errors" were counted as the number of times the rat made sideways head movements at the choice point before taking one or the other pathway. Totals were taken for Trials 5 to 15, inclusive.

---

[5]The failure of disruption to occur would not, however, disprove the hypothesis, since plateau performances might be relatively independent of the specific cues used in learning. Tryon (17, 19) has shown that complete and sudden removal of visual cues during the plateau period of learning the *alley* maze has no effect on individual differences in errors and only a very slight effect on speed of running. Nevertheless, it is possible that on the elevated maze, where visual cues are more prominent, both the learning and plateau performances depend to a greater extent upon visual control. If so, differential sensory preferences might have a greater effect on the relative scores of Brights and Dulls in this apparatus.

[6]A light of low intensity was used with the filter, which transmitted essentially no light below 580 millimicrons. Although no conclusive proof exists that rats are unable to see light above this point, their behavior during the experiment appeared to justify assuming that they could not. In Figure 6, where the total time curves have been reproduced, it is evident that a marked disruption occurs with introduction of the red light at Trial 13, whereas no change can be detected at Trial 17 between the red light condition and total darkness.

I. Six-Unit Discrimination

This apparatus consisted of six consecutive discrimination units, each of which was essentially a modified Lashley jumping stand containing two possible openings for the rat to pass through. The openings were covered by light-weight gates, hinged at the tops and fastened (in the case of the "*wrong*" gates) *at the bottoms*. White, black, or neutral color cards were pasted on the fronts of these. Passing from one unit to the next, the rat jumped a gap of six inches to a narrow platform in front of the two gates. A vertical partition separated the gates, forcing the rat to jump back across the gap when he had chosen the wrong side. The platform in back of each unit was narrowed towards the center, making it necessary for the next choice to be made from a central position.

The rats were started from sliding cages (rather than being placed on the initial jumping platform by hand) and were run through the six units to sliding cages containing food at the end. Fourteen preliminary trials were given for training in jumping and in operating the gates, during which neutral gray cards were pasted to all gates and the right-left sequence of correct sides was randomly varied.

The test problem was set up as an ordinary black-white discrimination-learning series, except that the right-left positions of black and white gates were not varied. The white-correct gates were arranged in the order, *L-R-R-L-L-R*, and were kept in that order throughout the series. This made it possible for the problem to be learned either in terms of the constant visual cue or as a constant spatial-temporal pattern.

The rats were run two trials per day for 24 trials, at the end of which all groups had reached an average level approaching zero errors. This was *the learning series, during which the following measures were taken for inclusion in the 30 test variables:*

> 27. *Errors:* A maximum of one error could be made in each unit; scores were the total errors for 24 trials.
>
> 28. *VTE:* One "vicarious trial" was *counted for each head movement* in the direction of either gate.
>
> 29. *Starting Time:* Time in seconds spent on the starting platform *before jumping to the first* unit.
>
> 30. *Time:* The remaining time in seconds up to the point of reaching the end cages.

On the 25th trial, the right-left sequence of white-correct gates was reversed to the new order, *R-L-L-R-R-L*, as an exploratory comparison of Brights and Dulls with respect to relative strengths of the spatial and visual

cues. Four trials, including the 25th, were run under these conditions. The results of this test may be mentioned here briefly as of some possible interest for future study, but they were somewhat ambiguous and are not sufficiently reliable to be included with the other data. The Brights' errors increased nearly to "chance" on the 25th trial, whereas Dulls were only slightly disrupted. On the surface, this result is similar to Krechevsky's previous finding with regard to spatial and visual "hypotheses," but the other measures which were taken at the same time indicate that in this case a somewhat different principle may be involved: (*a*) As indicated by their starting times and *VTE*'s[7] both the Brights and Dulls were noticeably disrupted at the *first* unit, while (*b*) Brights' errors were reduced immediately on the following trials. These observations indicate that both groups learned both cues, and that the differentiation here is therefore one of *preference* for spatial or visual—after learning has occurred—rather than of contrast in processes which contribute to the rats' learning. Thus, the present results provide no direct evidence either for or against the theory that a "spatial" tendency underlies the Brights' superior alley maze performances. The experiment should be repeated with larger groups and with different types of variations being introduced at the crucial trial.

[7] *VTE*'s (vicarious trials and errors) are ordinarily considered to be a "cognitive" rather than an emotional behavior (cf. Ref. 4, p. 19). But in the case of this particular trial on the discrimination apparatus, the increase in back-and-forth head wags was clearly a sign of disruption.

III. STATISTICAL ANALYSIS

A. Standardization of Test Measures

The raw scores for all rats in each of the 30 tests are listed in Table 3. To convert these into units of a common scale, using *MxM*'s as the standardization group, a table was prepared (Table 4) by means of which the *MxM* raw scores in each variable, when ranked, could be converted directly to normalized values on a scale ranging from 0 to 100. In constructing the table, ranks were first transformed into percentile ranks and then into their sigma score equivalents as found in a table of normal curve areas. The sigma scores were then converted to final scale values by means of the formula, $S = 20z + 50$.

The final scale had, therefore, a mean of 50 and a standard deviation of 20. Because of the small number of rats in the *MxM* group, their scores actually covered only the range between 13 and 87 (corresponding to −1.83 and +1.83 sigma units, respectively), the full scale from 0 to 100 having been allowed in order to cover a few instances in which individual Brights' and Dulls' raw scores were definitely above or below those of the extreme *MxM*'s. Ideally, the limiting scores would be set farther out on the scale in terms of sigma units, since it is reasonable to expect that members of the pure strains might deviate as much as three standard deviations in some variables. By the same token, a much larger sample of *MxM*'s would be highly desirable, both to eliminate the need for extrapolation and to increase the reliability of obtained points within the distributions. It was found, however, that the majority of Brights and Dulls were within the *MxM* ranges on most of the tests and it was therefore considered inadvisable to extend the limits beyond those chosen.[8]

In order to transform individual Brights' and Dulls' raw scores into the new scale values, a graph was made for each variable on which *MxM*'s normalized scores were plotted as a function of their raw scores. A smooth curve was drawn through the plotted points and scores for the experimental groups were read directly from these graphs. Since approximations were necessarily involved in the smoothing process, the reading of normalized scale values could not be considered highly accurate. Hence, the Brights' and Dulls' scores were read only to the nearest multiple of five units. The results are listed in Table 5.

[8] *Unreliability of the scales caused by random variation in means and sigmas of the small MxM sample obviously does have some effect upon the obtained patterns of scores for Brights and Dulls, but the nature of the effect is difficult to estimate.* Considering the probable amounts of sampling error, the influence upon relative values of the correlations within and between strains should be negligible.

TABLE 3

TABLE OF RAW SCORES FOR BRIGHTS, DULLS, AND *MxM*'S IN THE 30 TEST VARIABLES

Rat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
B- 6	16	1	4	0	10	16	209	6	93	60	134	198	267	118	20	160	76	2212	111	36	4	565	300	353	13	21	61	218	5	12
B- 5	14	2	9	4	24	47	309	11	76	55	130	285	255	150	21	150	124	7462	141	20	2	413	298	216	16	20	40	255	7	18
B-10	8	1	11	3	114	27	224	17	98	63	130	45	287	90	14	140	74	2737	28	18	0	616	228	176	14	8	27	232	4	16
B-12	10	7	27	23	69	16	179	7	56	132	191	112	270	140	21	165	87	11301	138	8	10	468	285	178	11	9	28	280	5	13
B-16	15	7	24	4	59	23	206	2	40	72	153	34	301	115	9	175	97	702	153	26	25	661	404	370	17	19	25	279	4	12
B-18	17	7	31	10	100	20	182	2	30	72	120	31	236	115	19	160	83	4974	72	7	10	590	411	272	14	28	38	220	3	23
B-26	8	1	4	0	55	17	160	9	44	75	133	30	625	156	15	115	74	3572	195	11	20	440	241	305	18	17	50	223	7	11
B-27	9	2	14	31	87	16	146	14	71	75	126	26	181	106	22	150	82	3539	207	24	15	540	279	272	15	21	42	238	6	24
B-28	4	1	28	13	76	14	145	5	27	35	122	172	247	128	11	145	69	5881	449	25	10	546	272	189	12	12	44	211	2	11
B-30	8	0	17	23	112	28	175	10	58	85	140	34	302	116	19	150	145	2350	96	18	1	416	290	209	24	3	39	275	23	44
B- 3	0	1	34	0	136	115	540	12	102	53	108	30	159	155	17	150	52	33015	59	14	2	387	225	344	24	15	31	306	22	34
B- 6	1	0	35	4	140	124	656	14	212	29	102	21	115	240	23	290	161	5755	47	43	7	345	571	411	47	58	21	400	30	60
B- 9	0	1	32	3	186	99	493	17	142	34	109	112	154	220	25	175	80	27623	62	22	7	485	443	331	19	20	38	287	30	55
B-10	0	0	32	14	139	122	456	14	94	21	165	51	118	210	24	235	86	10580	105	24	6	415	302	256	27	11	46	276	12	34
B-12	0	1	23	9	77	124	829	15	160	27	100	35	147	150	20	170	74	30485	44	54	5	549	353	358	17	37	31	244	6	19
B-15	7	0	9	4	28	89	810	9	177	34	97	38	113	160	17	220	190	6560	133	24	5	503	464	392	45	50	19	294	20	28
B-18	0	0	31	3	97	149	517	16	97	11	86	33	125	135	28	250	161	13029	56	19	4	421	543	373	32	50	36	381	32	28
B-22	0	0	28	14	96	120	598	16	103	31	107	28	194	170	13	155	128	15430	61	39	11	590	353	296	35	18	46	243	8	34
B-24	14	0	21	23	126	147	698	19	146	57	108	59	144	160	17	130	109	10315	63	18	4	308	281	346	24	9	60	194	8	24
B-23	0	2	49	3	164	135	596	12	176	14	111	26	152	265	27	265	170	8122	59	14	4	308	349	330	23	11	23	295	12	27
BR 5	17	9	5	2	14	108	350	10	48	19	90	26	180	120	19	145	110	20555	53	23	6	370	283	197	81	22	48	245	6	27
BR 9	5	4	21	4	137	43	175	10	51	52	107	40	193	105	13	135	78	50629	28	44	28	544	324	253	18	50	32	196	9	8
BR11	8	1	15	10	78	42	285	10	54	19	116	27	154	125	21	175	111	15589	74	32	32	362	329	308	21	25	46	264	7	16
BR15	2	3	23	2	159	74	350	20	180	20	159	20	153	150	8	140	150	16000	83	45	9	379	400	231	20	31	27	306	12	20
BR17	0	1	35	4	159	92	269	10	68	19	97	35	217	130	28	160	145	36693	41	30	20	526	353	266	9	33	43	221	4	11
BR22	2	6	25	14	186	132	384	11	74	46	124	50	149	120	25	225	114	28272	35	24	22	361	301	329	24	22	40	230	6	24
BR26	0	1	19	8	214	32	152	6	34	35	95	18	134	100	15	140	180	5882	280	38	5	437	302	192	18	6	39	240	14	28
BR30	0	5	24	12	191	33	192	15	109	30	120	21	99	95	7	125	124	44736	55	54	14	566	358	227	23	54	38	197	5	15
BR43	0	0	29	0	172	56	251	10	59	20	142	153	225	105	23	110	92	2747	108	52	28	401	265	334	13	13	58	230	9	27
BR37	0	2	21	6	200	55	279	19	182	57	100	19	186	155	11	130	88	10762	118	24	2	272	277	290	31	12	35	210	3	11
BR41	10	10	34	18	70	65	274	8	82	63	106	25	169	210	19	180	137	8066	55	36	18	347	398	400	35	28	42	237	6	19
BR45	0	1	11	7	170	102	350	17	77	27	99	41	186	155	22	135	127	25859	55	10	17	669	371	324	47	24	39	272	13	36
BR48	0	1	35	2	192	89	169	17	103	23	127	34	179	100	17	115	64	23074	62	49	21	407	275	177	21	11	37	200	6	14
BR54	4	2	56	8	183	80	240	14	65	60	110	72	192	115	23	150	95	14752	68	32	25	453	275	238	18	5	47	208	6	21
BR56	7	8	16	14	146	44	212	19	47	49	105	66	127	110	14	125	99	16956	46	39	19	389	221	173	13	10	41	175	2	10

TABLE 4

TABLE FOR CONVERTING RANKED SCORES OF THE *MxM* GROUP INTO FINAL SCALE VALUES

Rank	% Rank	z	S
1	96.67	1.83	87
2	90.00	1.28	76
3	83.33	0.97	69
4	76.67	0.73	65
5	70.00	0.53	61
6	63.33	0.34	57
7	56.67	0.17	53
8	50.00	0.00	50
9	43.33	—0.17	47
10	36.67	—0.34	43
11	30.00	—0.53	39
12	23.33	—0.73	35
13	16.67	—0.97	31
14	10.00	—1.28	24
15	3.33	—1.83	13

B. DISTRIBUTIONS OF SCORES FOR INDIVIDUAL TESTS

The data of Table 5 are presented graphically in Figure 4, which shows histograms drawn for each of the 30 test measures. The baseline of each figure represents the possible range of values from 0 to 100, extending from left to right. Individual Bright rats are represented as crosses above the line and the Dulls as circles below.

Because of the method by which the scores were obtained, it is to be assumed that *MxM* rats distribute normally on each of the 30 test measures, having means of 50 throughout. This background comparison is taken to be implicit in all of the Brights' and Dulls' scaled scores as they are discussed in the following sections.

It may be noticed in comparing Table 5 with the histograms that the direction of "high" and "low" has been reversed in many cases from that of the original scores. For example, in the case of Variable 6 (errors in the Tryon Maze), a high scale value represents a low number of errors, and in Variable 7, *the high scale value represents a low total time score.* All time and error scores were treated in this way in conformity with the accepted notion of what constitutes "good" and "poor" learning performances. Some of the other variables, however—e.g., Variable 19 (percentage pre-feeding activity) and Variable 28 (*I'TE*'s in discrimination learning)—required an arbitrary decision. In such cases, the direction chosen to be "high" was that which made for positive correlation within the *MxM* group with other measures taken in the given apparatus. Since time and error scores constituted

TABLE 5

TABLE OF SCALED SCORES FOR BRIGHTS AND DULLS

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Rat #
15	65	15	85	15	100	60	85	35	25	10	10	5	55	45	G- 6
20	50	20	55	20	50	15	60	50	35	10	0	5	30	40	G- 5
30	65	25	60	35	100	55	30	35	15	10	35	5	80	60	G-10
25	30	60	10	25	100	70	75	60	0	0	20	5	35	40	G-12
15	30	55	55	30	100	60	95	80	5	5	50	0	55	75	G-16
15	30	65	40	30	100	85	95	90	5	10	50	10	55	45	G-18
30	65	20	85	25	100	85	70	75	5	10	50	0	35	60	G-24
30	50	30	0	25	100	90	45	50	5	15	55	45	65	65	G-27
40	85	50	30	25	100	90	95	90	45	25	10	10	45	70	G-28
30	85	40	10	35	100	75	65	55	5	10	50	0	55	70	G-30
55	65	70	85	40	15	0	55	35	45	50	50	70	25	50	B- 3
55	85	70	55	40	20	0	45	10	50	60	65	80	5	35	D- 6
65	65	70	60	55	25	0	30	25	45	45	20	70	10	20	B- 8
65	85	70	30	40	20	0	45	35	60	55	35	80	15	30	B-10
65	65	50	55	30	20	0	40	20	55	60	45	65	30	45	B-12
35	85	25	55	20	30	0	70	20	45	70	45	80	25	50	B-15
65	85	65	60	30	10	0	35	35	95	90	50	75	35	20	D-18
65	85	80	30	30	20	10	35	35	50	60	55	60	20	65	B-22
20	85	45	10	35	10	0	25	25	30	50	30	65	25	50	B-24
65	50	80	60	50	15	0	25	20	90	45	55	65	0	20	B-26

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Rat #
35	70	10	30	45	75	10	60	25	75	50	25	45	55	65	G- 6
40	45	30	20	75	80	35	50	60	60	50	50	65	40	50	G- 5
50	70	15	90	75	90	15	85	80	70	75	85	55	60	55	G-10
35	65	35	20	90	65	20	60	75	80	70	85	55	70	60	G-12
30	65	5	15	65	35	5	20	20	60	50	90	75	60	65	G-16
35	85	20	40	90	65	10	15	45	70	45	55	45	70	40	G-18
75	70	15	10	85	45	25	40	40	65	55	80	35	40	65	G-24
40	85	15	5	70	60	10	60	45	70	50	45	40	45	35	G-27
45	75	25	0	85	65	70	65	70	75	60	40	40	85	65	G-28
40	30	10	35	75	90	35	55	65	40	90	65	75	10	10	G-30
40	85	65	70	80	80	45	45	25	40	55	80	85	10	15	B- 3
0	5	25	60	30	70	70	0	10	25	10	95	100	0	0	B- 6
30	70	65	60	70	70	20	10	15	55	50	55	85	0	5	B- 8
10	60	35	30	70	70	35	60	60	40	65	35	75	30	15	B-10
30	70	65	65	10	75	10	35	20	60	10	80	60	45	60	D-12
15	5	25	20	70	75	15	10	15	25	0	100	95	10	30	B-15
10	10	40	55	75	75	30	0	20	35	40	65	95	0	30	B-18
40	40	50	40	40	65	45	35	40	30	55	35	60	35	15	B-22
60	50	35	50	80	75	80	60	25	40	70	10	30	35	35	D-24
5	10	30	50	80	75	75	35	15	45	65	95	95	30	30	B-26

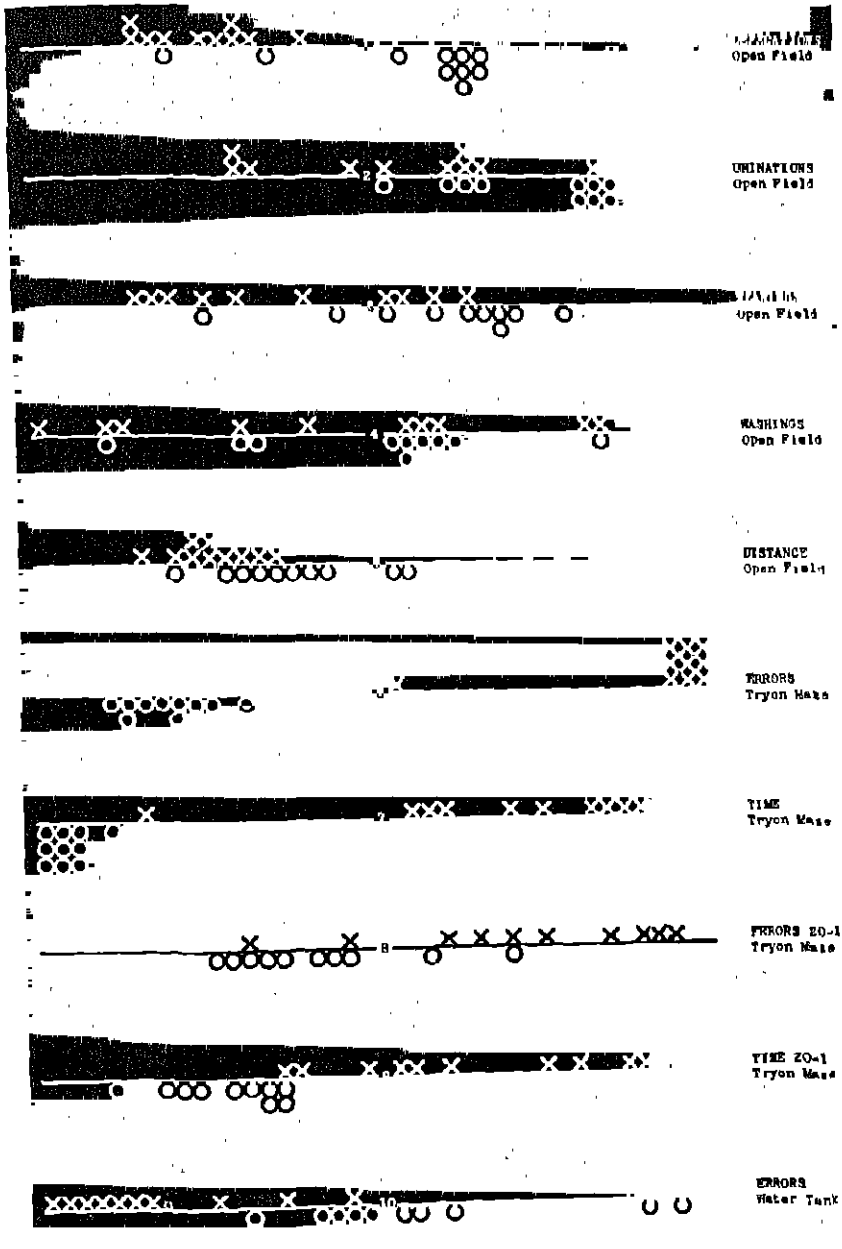


FIGURE 4a

HISTOGRAMS OF INDIVIDUAL VARIABLES: VARIABLES 1-10

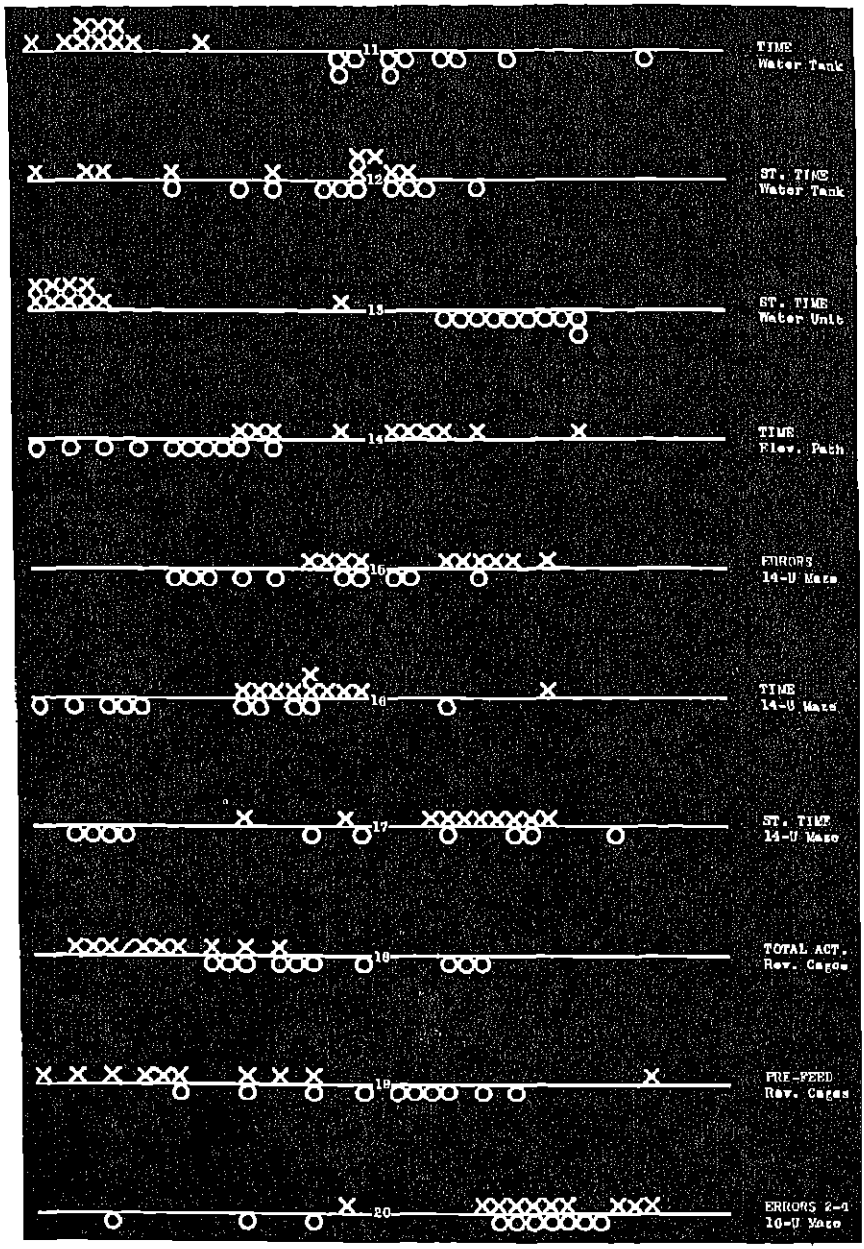


FIGURE 4*b*
HISTOGRAMS OF INDIVIDUAL VARIABLES: VARIABLES 11-20

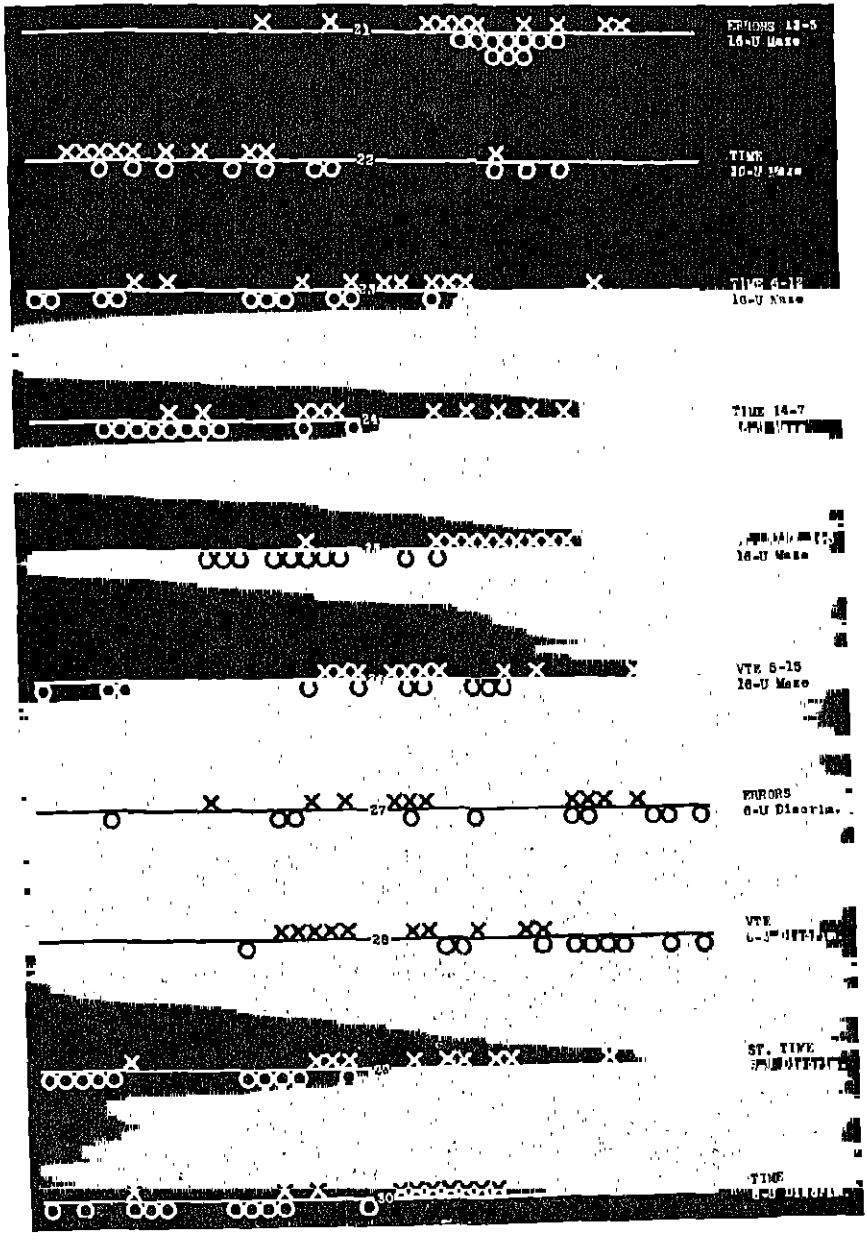


FIGURE 4c

Histograms of Individual Variables: Variables 21-30

most of the measures taken, only four of the 30 test variables were *not* reflected in the process of scaling. These non-reflected variables are listed below:

3. *Rearings:* Open Field.
5. *Distance:* Open Field.
18. *Total Activity:* Revolving Cages.
28. *VTE:* 6-Unit Discrimination.

C. Intercorrelations Among Rats

It is generally true of all methods of multiple factor analysis that the variables which have high loadings in any given factor (i.e., which form a cluster in common factor space) are variables which correlate highly with one another and low with other variables outside the cluster. Hence, if it can be shown in the present case that individual Bright rats correlate highly with one another and low with individual Dulls, and that Dulls also correlate highly among themselves, it may be assumed that Brights and Dulls would form two separate clusters in common factor space if a complete factor analysis were performed on the data. Since the question of principal interest in the present experiment involves the possible existence of only these two types, and since members of the types are known in advance, it was considered economical to test the hypothesis directly by simply finding the average intercorrelations within and between groups.

Accordingly, averages were computed for (*a*) the correlations of Bright rats with other Brights, (*b*) of Dulls with other Dulls, and (*c*) of Brights with Dulls. In order to provide a means of judging the representativeness of group averages, mean correlations were also computed separately for each rat.[9] The results are shown in Table 6.

With reference to the hypothesis that Brights and Dulls are different behavior "types," it may be seen by referring to Table 6 that the obtained correlations definitely indicate a positive answer. The fact that the negative correlation *between* groups (—.19) is much smaller in magnitude than the average positive correlation *within* each group (+.56) indicates, further, that the two strains are not merely opposites but are comparatively unrelated. This finding is inconsistent with the conventional assumption of simple, quantitative differences among individuals: it implies that different *patterns* of components must be postulated to underlie maze-brightness and maze-dullness.

[9]The 190 possible intercorrelations among individual rats were not computed. Instead, the values listed in Table 6 were obtained by means of formulae for average intercorrelations, based upon scores which were first standardized for each rat with reference to his distribution of scaled scores in the N test variables. No approximations were necessary; the obtained values represent true arithmetic means of the intercorrelations.

TABLE 6
AVERAGE INTERCORRELATIONS

Brights	Mean r with Brights	Mean r with Dulls
*A. Averages for Individual Rats**		
G- 6	.592	—.198
G- 5	.559	.079
G-10	.522	—.152
G-12	.663	—.233
G-16	.606	—.195
G-18	.629	—.260
G-24	.624	—.268
G-28	.568	—.353
G-30	.540	—.025
Mean Bright	.587	—.188
Dulls		
B- 3	—.140	.609
B- 6	—.270	.602
B- 8	—.155	.597
B-10	—.136	.591
B-12	—.244	.431
B-15	—.016	.502
B-18	—.300	.619
B-22	—.265	.555
B-24	—.112	.273
B-26	—.238	.554
Mean Dulls	—.188	.533

B.—Averages for the Groups

Group	No of rats	No. of inter-r's	Ave. inter-r
Bright vs. Bright	10	45	.587
Dull vs. Dull	10	45	.533
Bright vs. Dull	20	100	—.188

*The mean r of each rat with his own group represents the average of nine correlations; that with the opposite group, the average of 10.

Since satisfactory formulae were not available for computing standard errors of average intercorrelations, precise estimates could not be made of their statistical significance. In view of the non-random character of the sampling, standard errors which assume simple sampling would probably be too high. That is, since the Bright and Dull groups had been selected by taking one member from each litter of the given generation, they were probably more representative of their respective populations than estimates based on conditions of random sampling would indicate. Nevertheless, a fair evaluation of the reliability of group averages may be made by consulting the values in Table 6 for individual rats. Whereas the correlations of

individual Brights with other members of their own group range from +.52 to +.66, their correlations with members of the Dull group range from —.35 to +.08. Similarly, the individual Dulls correlate from +.27 to +.62 with other Dulls, but from —.30 to —.02 with Brights. In comparison with the differences between groups, the ranges of intra-group variation are very small. Since these figures indicate that members of the two strains can be distinguished with perfect accuracy on the basis of their score patterns alone, there can be little question that the group differences are significant statistically.

In evaluating the figures of Table 6 it should be kept in mind that the obtained values were subject to influence by both (*a*) unreliability of the test measures and (*b*) variable factors of previous training and experience. As regards the problem of the extent to which innate determiners are responsible for differences between Brights and Dulls in their trait organization, both of these secondary types of influence are spurious, and both would operate in such a way as to lower the obtained correlations. If, therefore, the "true" values could be computed, it is reasonable to suppose that they would show both a closer similarity of members within each strain and a sharper difference between the strains.[10]

[10]Estimates might be made of these true values by either of two methods, neither of which was used for the present report because its value did not seem to justify the required labor to compute: (*a*) Attenuation by factors of measurement could be corrected for if reliability coefficients were first computed for each rat. The number of these coefficients, 20, is not great, but to compute them would also require the preliminary scaling of 60 split-half score distributions. (*b*) If a complete inverted-factor analysis were made, the relationships when expressed in terms of "common factor" correlations should approximate the true degrees of hereditary resemblance, since the effects of both unreliability and specificity (which, presumably, would be increased by environmentally-produced differences among rats) would be removed.

IV. PSYCHOLOGICAL ANALYSIS OF BEHAVIOR PATTERNS

A. Graph of Bright and Dull Medians

In order to obtain a simplified picture of the Brights' and Dulls' performances on *all* of the variables, a chart was prepared which is reproduced in Figure 5. The chart is constructed in the form of a scatter diagram, showing the position of each test measure with respect to the median scores of the two groups of rats. Scale units of the ordinate, representing Dulls, and of the abscissa, representing Brights, are in terms of the final scaled scores, which range from 0 to 100. Each test measure is plotted as a point whose coördinates are the medium scores of Brights and of Dulls, respectively, in that test. Thus, for example, the coördinates of Variable 6 are 100 for Brights and 25 for Dulls, these two values having been determined as the respective medians in that variable (cf. histograms of Variable 6 in Figure 4).

In general, since the score of 50 indicates the constant position of the median *MxM*, all points in the lower right hand quadrant are variables in which the median Bright is above, and the median Dull below, the average *MxM*. The reverse is true for points in Quadrant II (upper left), while in Quadrants I (upper right) and III (lower left) both groups are above and below average, respectively.

The scattering of points indicates a slight negative correlation between the median Bright and median Dull, as might have been expected in view of the previously obtained mean correlation of —.19 between individual members of the two groups.

The succeeding analysis of group differences is based upon these median profiles. As a measure of the extent to which they may be considered representative of the profiles of individual rats, each rat's score distribution was correlated with the median for his group. It was found that Brights correlated +.78, on the average, with the median Bright profile and Dulls +.76 with the median Dull profile. Although far from indicating perfect homogeneity, these figures are sufficiently high to justify accepting the median profiles as characteristic of their respective groups.

B. Generality of Brightness and Dullness

By analyzing the scatter of variables in Figure 5 it is possible to note the degree to which obtained scores are consistent with various hypotheses concerning the nature of specific behavior differences.

A consideration of primary interest concerns the extent to which there is a general superiority of Brights over Dulls. This does appear to be the

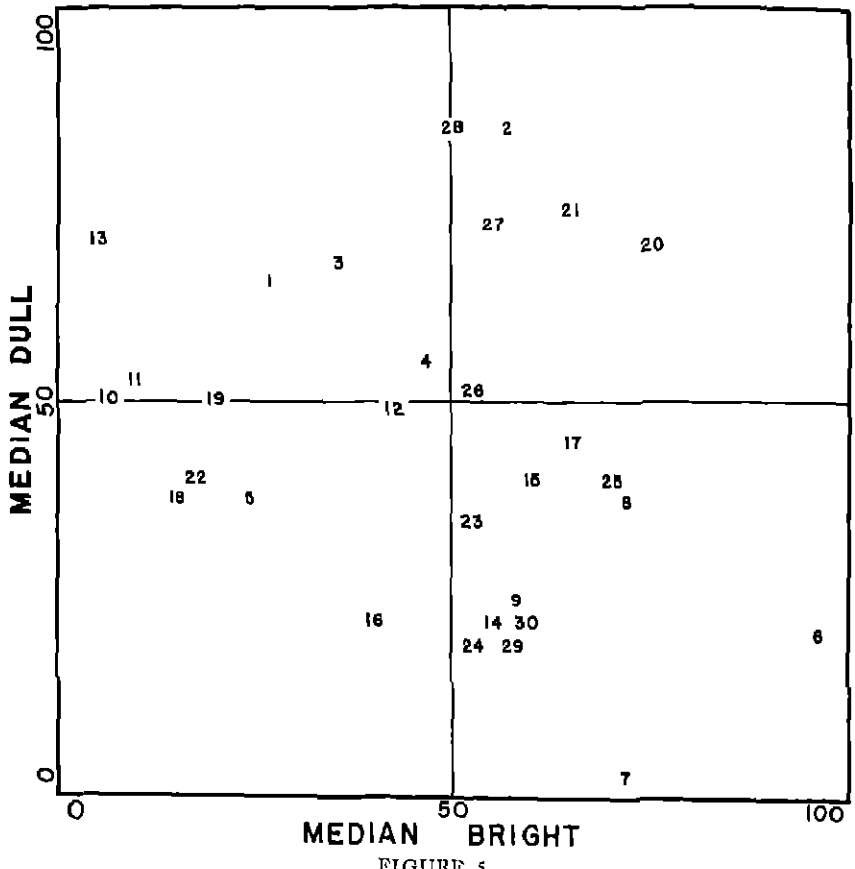


FIGURE 5

Scatter diagram showing the relative positions of Bright and Dull medians in all variables. Each numbered point in the plot represents a test variable, the locus of which is determined, on the abscissa, by the median score for Brights on that variable, and on the ordinate by the median score for Dulls.

case to a slight degree, as it may be observed that the variables lie to the right of the center vertical line, and below the center horizontal line, in more than half of the cases, indicating that Brights are "high" and Dulls are "low" in a majority of the measures. The difference, however, is too small to be considered reliable. It is significant that Dulls were actually superior to *MxM*'s in 13 of the measures (i.e., Variables 1, 2, 3, 4, 10, 11, 13, 19, 20, 21, 26, 27, 28) while Brights were below average in 12 (Variables 1, 3,

4, 5, 10, 11, 12, 13, 16, 18, 19, 22), a finding which obviously indicates that Dulls are not *universally* inferior rats and that Brights are not universally superior.

Of particular interest however, are the measures of learning. Variables 6 (Tryon Maze), 10 (Water Tank), 15 (14-Unit Maze), 20 (16-Unit Maze), and 27 (6-Unit Discrimination), all being measures of error, are the five measures which best represent the rats' learning performances under a variety of conditions. Referring to the positions of these variables in Figure 5, it may be seen that the Brights are above average in 6, 15, 20, and 27, and very low in 10, while the Dulls are low in 6 and 15, average in 10, and definitely superior in 20 and 27. The Dulls, then, are either equal to or better than Brights in three of these five measures. In view of this result, it is apparent that the performances of the two groups on any particular apparatus are not to be accounted for in terms of a difference in general learning ability.

C. Learning in Food-Reward Situations

Of the five learning variables discussed above, the median Bright score was found to be below average only in Variable 10. This variable is also the only one of the series in which the reward employed was escape-from-water rather than food. When only those measures representing errors in food-reward situations are considered, it is found that the Brights made low error scores consistently. Thus in Figure 5*a*, where the food-reward error measures are graphed separately,[11] all of the points appear to the right of the center vertical line, indicating that the Brights were generally superior in eliminating blind alley entrances when the incentive was food. On the basis of this finding, one is tempted to suggest that the error scores are determined primarily by degrees of food motivation. However, scores for the median Dull are distributed throughout both Quadrants I and IV, and such an interpretation, if applied to all three groups of rats, would imply that Dulls were approximately equal to *MxM*'s in food drive. In view of the previous indications that the groups are differently organized, the possibility is still not precluded that Brights' superiority is due to the strength of their food drive. However, conclusions regarding this point should be reserved until further evidence is considered.

[11]Variables 8 and 21 represent plateau-trial performances; they were therefore not included in the group discussed in the preceding paragraph.

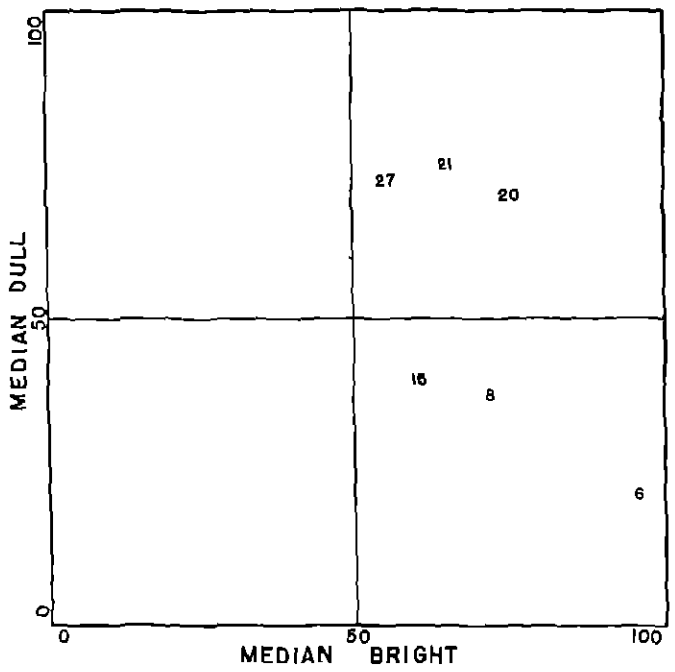


FIGURE 5a

ERROR MEASURES: FOOD PROBLEMS

D. MOTIVATION TO RUN FOR FOOD

The measures of starting time and total time of running relate more directly to motivational levels of the groups. These variables are plotted separately in Figure 5*b*. It may be noted by referring to the figure that whereas the points are scattered on both sides of the center vertical line, they are all in the lower half of the diagram, indicating that Dulls were universally low in these measures. It was observed on numerous occasions during the experiment that Dulls were typically apathetic towards food. They were slower to begin eating at the food box than were Brights and *MxM*'s and often failed to eat all of their daily ration.[12] On the whole, the evidence indicates that Dulls as a group are characteristically below average in food drive.

As regards Brights and *MxM*'s, no consistent differences were observed

[12]In this connection it may be recalled that the Dulls' weights are also lower.

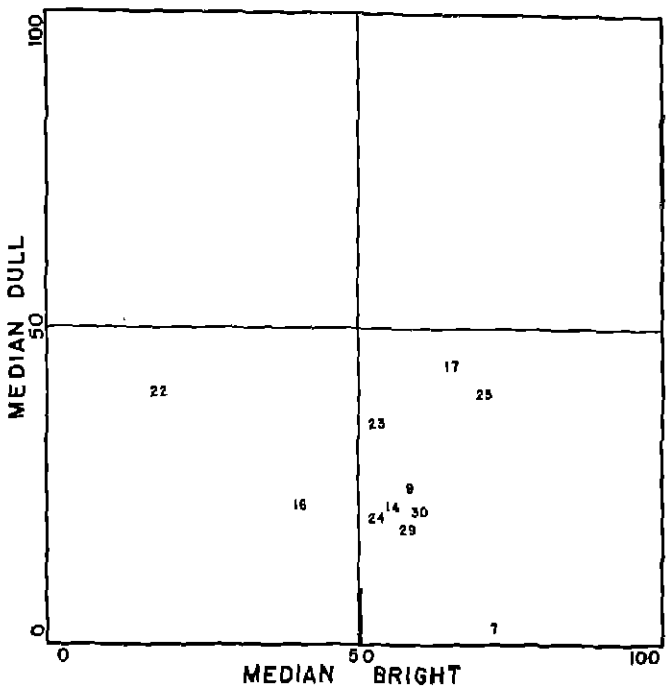


FIGURE 5*b*
Time Measures: Food Problems

either in running speed or in eating behavior. The points in Figure 5*b* indicate that Brights' time scores were somewhat lower on the average (i.e., "better") than those of *MxM*'s, but these scores were influenced by various factors other than speed of running. For example, the marked influence of errors may be seen by comparing the positions of Variables 6 and 7, the measures of errors and time, respectively, on the Tryon Maze. Variable 7 would have appeared much closer to the mean *MxM* if corrected for number of errors. Insofar as strengths of food drive can be inferred from these sources of evidence, then, it may be supposed that Brights are no more strongly motivated by the food incentive than are average rats.[13]

[13]That is, in an absolute sense. As discussed in a later section, there are indications that food is stronger than other motivating factors for Brights and therefore is probably a more significant factor in their behavior as a whole than for other rats.

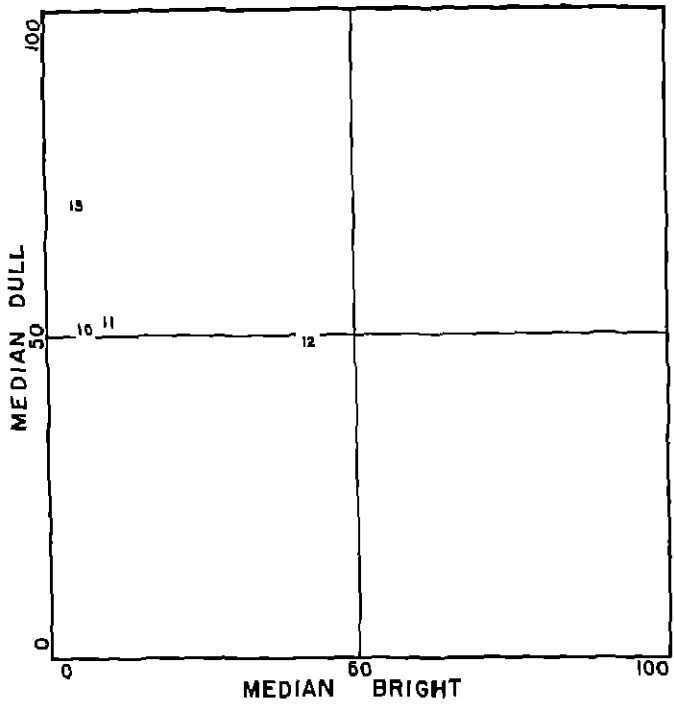


FIGURE 5*c*
TIME AND ERROR MEASURES: WATER PROBLEMS

E. ESCAPE-FROM-WATER MOTIVATION

In Figure 5*c* are shown the time and error measures which were taken for the Water Tank and the Water Unit. In contrast with their performances in food-reward problems, the Brights proved to be consistently poor in both time and errors, while Dulls were average or better. Although further study is needed in order to evaluate these results fully in relation to the rats' maze behavior, they indicate clearly that motivational differences of a complex character exist between the two strains. The fact that Brights' scores are correspondingly low in revolving cage activity (see discussion below) suggests that their performances here may be related to their low level of spontaneous energy.

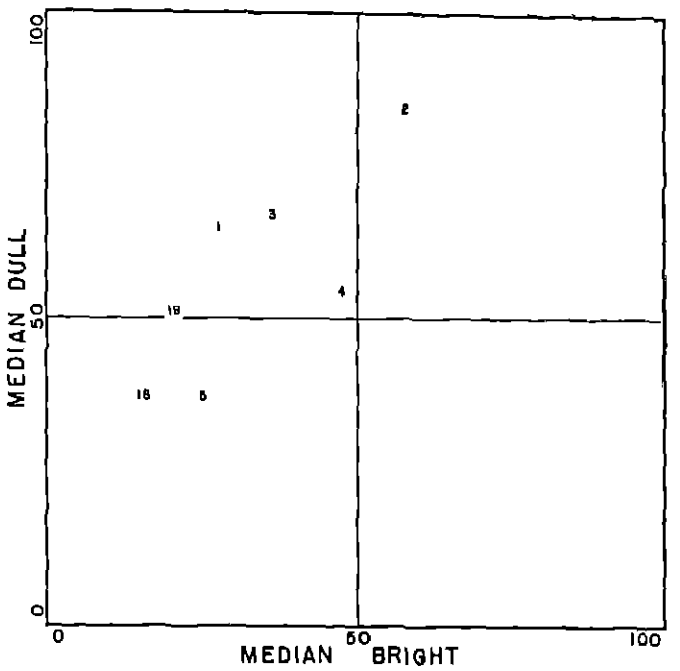


FIGURE 5*d*
OPEN FIELD AND ACTIVITY MEASURES

F. ACTIVITY AND EMOTIONALITY

A fourth selection of measures is shown in Figure 5d. These are Variables 1 to 5, inclusive, which were taken in the Open Field apparatus, and Variables 18 and 19 representing Revolving Cage activity.

In Revolving Cage activity, both pure strains were lower than the *MxM* group. Brights, however, were definitely less active than Dulls and, as may be seen by referring to the histogram of Variable 18 in Figure 4, even the most *active of the individual* Brights was well below the average *MxM*. The fact that the "exploratory distance" measure taken in the Open Field (Variable 5) showed such similar results suggests *that both of these measures* reflect the rats' general activity levels.

The Brights' position in Variable 19 is particularly significant, since this variable provides a direct indication of the relative strengths of their food and spontaneous activity drives in comparison with Dulls and *MxM*'s. As

there was no good criterion for evaluating the percentage scores in terms of high and low, the variable might have been placed on the opposite side of the graph with equal justification. However, its position indicates essentially that what little activity the Brights showed was due to hunger. Inspection of the raw scores reveals that the Brights were 60 per cent more active during the two hours before feeding than for the entire remainder of the daily period, while for Dulls and *MxM*'s the bulk of activity occurred at night. Thus, since pre-feeding scores are included in Variable 18, an even greater discrepancy exists in the spontaneous activity levels of the three strains than this total activity measure indicates.

The Open Field measures tend to scatter rather widely over the upper left section of the graph, but their position as a group signifies that Brights showed more of the emotional behaviors than average while Dulls, on the other hand, were more "stable." On the whole, this result agrees with evidence from various other sources. In subsequent running on some of the other types of apparatus, Brights were found to be characteristically stubborn during the initial trials, showing at the same time the teeth-chattering, face-washing, and other signs which are typical of emotional upset. This behavior was first noticed during preliminary training on the Elevated Path, in which Brights' time scores, although equal to those of *MxM*'s in later trials, were much higher than both Dulls' and *MxM*'s in the beginning. It also occurred on the two elevated mazes and on the 6-Unit Discrimination apparatus, each time being confined, however, to the early trials. It might be characterized, then, as a response to unfamiliar "open space" situations. Direct measures of the behavior were not obtained, but its influence on running times in each of the apparatuses was typically like that shown by the total time measures for the 16-Unit Maze. Referring to Figure 5*b*, it may be seen that Variable 22, times for the first two trials, deviates markedly to the left, while the positions of Variables 23 and 24, the plateau and final trial scores, respectively, are "normal." With regard to emotional responses of the Dulls, the only indications found in the present series of problems were incidental in nature. It was noticed during preliminary training on the 6-Unit Discrimination apparatus that the Dulls were very reluctant to step on a platform which was held by the experimenter (sometimes unsteadily) to aid their first crossings. A specific fear of unstable platforms seemed to be characteristic. In preparing to run final darkness trials of the 16-Unit maze, a plate switch was installed on the pathway at the "End" position (cf. Figure 3) for the rats to step on in passing. The plate was fixed to depress slightly with the rat's weight, turning on room

lights and thereby enabling total times to be recorded by stopwatch. Most of the rats stopped upon meeting the switch for the first time (at the end of Trial 14). The Dulls, however, were exceptionally upset by it and refused to cross until blocked from retracing by the experimenter. The effect carried over to the following trials (15 and 16), causing them to begin slowing up successively farther back into the maze on each trial. In Figure 6, which shows time curves for all three groups during the later trials, it is apparent that Dulls' total times were affected significantly.[14]

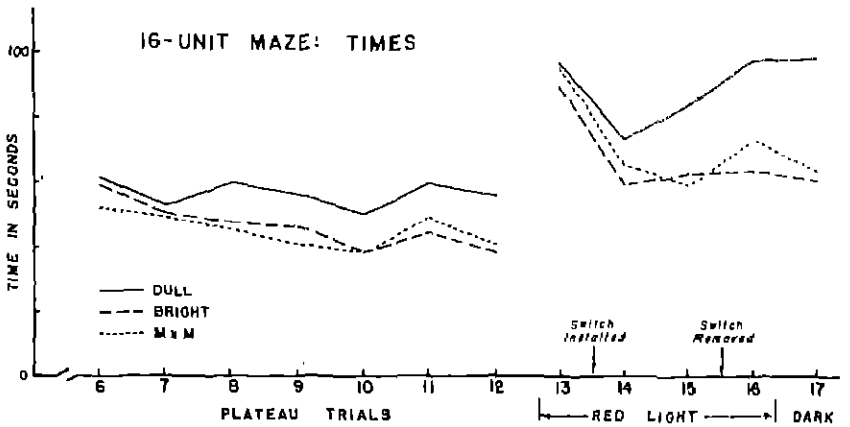


FIGURE 6
CURVES OF TOTAL TIMES FOR BRIGHTS, DULLS, AND *MxM*'S DURING PLATEAU AND FINAL DARKNESS TRIALS ON THE 16-UNIT ELEVATED MAZE

G. SUMMARY AND DISCUSSION OF THE ANALYSIS

The general findings of the preceding analysis may be summarized briefly as having indicated the following points:

1. Brights' scores on the 30 test measures tend to be slightly higher on the average than those of Dulls. However, this result depends upon selection of the particular tests which were included in the sample of test measures. With reference to the measures of errors in *learning*, where all types of the apparatus situations are considered, there is no evidence that Dulls are generally inferior.

2. As indicated by their running speeds and their behavior towards *food*, Dulls are below average in strength of food drive. No evidence was ob-

[14] The switch was removed just after Trial 15 but its absence was not discovered by the rats until the *end* of Trial 16.

tained which would permit differentiating between Brights and average rats with regard to food drive alone, but food appears to be more important for Brights in relation to other sources of motivation.

3. In problems involving escape-from-water motivation the Brights were consistently inferior, while Dulls were average or slightly better.

4. Both pure strains were below average in the Revolving Cage and Open Field measures of spontaneous activity. The Brights were particularly low, and showed the greatest *proportion* of their activity in the periods before feeding.

5. Different kinds of emotional responses appear to be characteristic of the two strains. There is evidence that Dulls have a specific fear of certain apparatus features, while Brights are initially timid in "open space" situations.

It remains to consider the possible relation of these various behaviors to the rats' hereditary maze-brightness and maze-dullness.

The finding in (1) above indicates that a "general intelligence" factor, if it exists at all, may be regarded as of little or no importance.[15] It was found, however, that the particular traits of each strain do appear to be "general" in the sense that they characterize the rats' behaviors in varying apparatus situations. From this, together with the intercorrelational evidence that Brights and Dulls are differently organized, it may be assumed that differences in the maze-learning ability represent differences in *patterns* of behavior traits rather than in *degrees* of any single psychological capacity.

1. *Brights*

The Brights' performances on all the learning problems where food was employed as an incentive may be taken to indicate that food drive is an important factor in their hereditary maze-brightness. It probably accounts to an appreciable extent for their superiority over Dulls. However, since the more direct indicators of motivation—running speed and behavior towards the food itself—showed no difference between the Brights and *MxM*'s, its importance seems to depend largely upon the relative strengths of other motivating factors.

Probably the most significant of the other factors is their exceptionally low spontaneous energy level. Both the Open Field distance measure and the 24-hour Revolving Cage score showed them to be lower than Dulls and *MxM*'s in general activity, while the percentage scores for pre-feeding sug-

[15] A similar result has been obtained consistently in previous correlational studies (7, 15, 22) and has been interpreted to mean that "specific factors" are largely responsible for individual differences in each type of apparatus.

gests that even this small amount should be attributed to hunger rather than to spontaneous energy. One might predict on logical grounds that rats having a motivational structure of this sort would show a "least distance" tendency in the alley maze, food-reward situation: being stimulated almost exclusively by the need for food and having little excess energy, they should be less distractible, should eliminate blind alleys more readily, and on the whole, should show behavior of a more "purposive" character than other rats.[16] In general, these expectations are consistent with Tryon's findings (19, p. 319) that Brights are relatively more controlled by the food-pointing, exit gradient, and higher direction-configuration components of the maze.

It may, however, be too easy to overestimate the importance of the food-activity balance in the Brights' total organization at the expense of other traits which have not been adequately identified. The "initial timidity" characteristic, for example, is potentially significant but its relation to maze error scores is not yet clear. The problem is somewhat complicated by apparent differences which have been observed as between the elevated and the alley maze situations. Tryon, Tryon, and Kuznets (20, 21) found that Brights were *less* emotional in their reactions to doors, curtains, alleys, and floors in the alley maze, but that in reaction to handling by the experimenter they showed more of the "avoidance" and "escape" responses. The latter finding suggests a similarity to the "open space" timidity shown in the present experiment, although handling was not always involved in this case. On the whole, the evidence indicates that the Brights' timidity is related to handling, to open-space and elevated types of apparatus, and to strangeness of the situation; but further studies are obviously needed before any precise definition of this trait should be attempted.

2. *Dulls*

In the measures of learning in food problems, Dulls showed no consistent tendency to be either high or low in errors, but their speeds of running were definitely below average. Considering the other indications that Dulls are generally low in food drive, the assumption is probably justified that low motivation accounts in part for their performances in the alley maze. However, since their errors in the other food problems were equal to those of Brights the importance of this factor should not be overemphasized. Either

[16]This hypothesis is not novel. Washburn, in an early experiment with white mice, compared error scores of the predominantly activity-driven mice with those of the predominantly hunger-driven and found that fewer errors were made by the latter group (23).

its influence upon errors is very slight, or other factors were operating in the elevated apparatus to compensate for its effect.

The indication that Dulls have a specific fear of unstable platforms should be studied further. It is conceivable that non-retrace gates and movable floors in the Tryon Maze would normally produce effects similar to those of the automatic light switch which was used during final trials of the 16-Unit Maze. Since non-retrace gates are always on the true path side of each choice point (necessarily), the effect of avoiding them would be an increase in errors as well as in total times. The implication may be that Dulls actually do learn the true path but that their learning, paradoxically, is manifested as entrances into, rather than as avoidance of, blind alleys. The various other facts which are known regarding the Dulls' psychological organization—namely, (*a*) their demonstrated ability to learn complex problems during the present experiment, (*b*) the indication of their unusually low food motivation, and (c) Tryon's finding that they show emotional reactions in the maze itself—are uniformly consistent with such a hypothesis.[17]

[17]Direct evidence in the Tryon Maze might be obtained by making systematic changes of movable floors and non-retrace gates, replacing them with devices of a different design. Dulls still might not perform as well as the Brights, but some improvement should be noticed if the hypothesis is correct.

V. GENERAL SUMMARY

An exploratory study of behavior organization in the Tryon strains of maze-bright and maze-dull rats was conducted with the purpose of determining the nature of psychological factors in their hereditary maze intelligence. As a result of preliminary studies in which the writer found that the two strains did not differ significantly in their abilities to learn an abstract spatial pattern, the hypothesis was suggested that Brights and Dulls are different psychological "types" of rats. The present experiment was designed to test this hypothesis and to identify the specific behavior tendencies which are thought to enter into each strain's maze-learning performance.

Thirty measures were obtained of learning, emotionality, activity, and other behaviors for samples of 10 Brights and 10 Dulls, and for a sample of 15 rats of a median strain. The latter group was taken as a reference group representing the normal rat population. The raw scores of rats in the two experimental groups were converted into standard scale values based upon scores of the reference group in each variable and these were used to obtain behavior profiles for members of the Bright and Dull strains.

As a modified form of the "inverted" factor analysis method, correlations were computed between the obtained score distributions for rats in order to determine the degree of homogeneity within each strain and the extent of similarity between strains. The mean correlation of Brights with Dulls was found to be low and negative (—.19), whereas members within each strain showed relatively high and positive correlations with one another (+.59, +.53). Although conclusions are conditioned to some degree by unreliability caused by the use of a small reference group, these results strongly confirm the type hypothesis.

No evidence was found that a difference exists between the Brights and Dulls in the learning capacity *per se*. A detailed study of the behavior profiles indicated that Brights are characteristically food-driven, economical of distance, low in motivation to escape from water, and timid in response to open spaces. Dulls are relatively disinterested in food, average or better in water motivation, and timid of mechanical apparatus features. It is concluded that brightness and dullness in the original Tryon Maze may be accounted for in large part by such motivational and emotional patterns. Although indications exist that the two strains may also be differentiated with reference to certain basic "cognitive" tendencies, the procedures followed in this experiment were not sufficiently analytical to indicate their nature. The need for further study is evident, both to relate the factors identified to the maze-learning performance in particular, and to discover other factors which may be present.

REFERENCES

1. Burt, C. The Factors of the Mind. New York: Macmillan, 1941.
2. Dunlap, J. W. The organization of learning and other traits in chickens. *Comp. Psychol. Monog.*, 1933, **9**, 1-55.
3. Flanagan, J. C. The Coöperative Achievement Tests: a bulletin reporting the basic principles and procedures used in the development of their system of scaled scores. New York: Coöperative Test Service, American Council on Education, 1939.
4. Geier, F. M., Levin, M., & Tolman, E. C. Individual differences in emotionality, hypothesis formation, vicarious trial and error, and visual discrimination learning in rats. *Comp. Psychol. Monog.*, 1941, **17**, 1-20.
5. Hall, C. S., & Ballachey, E. L. A study of the rat's behavior in a field: A contribution to method. *Univ. Calif. Publ. Psychol.*, 1932, **6**, 1-12
6. Krechevsky, I. Hereditary nature of "hypotheses." *J. Comp. Psychol.*, 1933, **16**, 99-116.
7. McCulloch, T. L. A study of the cognitive abilities of the white rat with special reference to Spearman's theory of two factors. *Contr. Psychol. Theor.*, 1935, **1**, 1-66.
8. Searle, L. V., & Brown, C. W. The effect of subcutaneous injections of benzedrine sulfate on the activity of white rats. *J. Exp. Psychol.*, 1938, **22**, 555-563.
9. Searle, L. V. Application of the cluster analysis technique to profile correlations of rats with brain lesions. *Psychol. Bull.*, 1939, **36**, 645.—Abstract.
10. Stephenson, W. Introduction to inverted-factor analysis, with some applications to studies in orexis. *J. Educ. Psychol.*, 1936, **27**, 353-367.
11. ———. The foundations of psychometry: four factor systems. *Psychometrika*, 1936, **1**, 195-209.
12. Thomson, G. H. The Factorial Analysis of Human Ability. New York: Houghton Mifflin, 1939.
13. Thorndike, R. L. Organization of behavior in the albino rat. *Genet. Psychol. Monog.*, 1935, **17**, 1-70.
14. Tolman, E. C., Tryon, R. C., & Jeffress, L. A. A self-recording maze with an automatic delivery table. *Univ. Calif. Publ.*, 1929, **4**, 99-112.
15. Tomilin, M. I., & Stone, C. P. Intercorrelations of measures of learning ability in the albino rat. *J. Comp. Psychol.*, 1934, **17**, 73-88.
16. Tryon, R. C. A theory of psychological components—an alternative to mathematical factors. *Psychol. Rev.*, 1935, **42**, 425-454.
17. ———. Studies in individual differences in maze ability: VI. Disproof of sensory components: experimental effects of stimulus variation. *J. Comp. Psychol.*, 1939, **28**, 361-415.
18. ———. Genetic differences in maze-learning ability in rats. *Thirty-Ninth Yearbook, Nat. Soc. Stud. Educ.*, Part 1, 1940, 111-119.
19. ———. Studies in individual differences in maze ability: VII. The specific components of maze ability, and a general theory of psychological components. *J. Comp. Psychol.*, 1940, **30**, 283-335.
20. Tryon, R. C., Tryon, C. M., & Kuzners, G. Studies in individual differences in maze ability: IX. Ratings of hiding, avoidance, escape, and vocalization responses. *J. Comp. Psychol.* 1941, **32**, 407-435.
21. ———. Studies in individual differences in maze ability: X. Ratings and other measures of initial emotional responses of rats to novel inanimate objects. *J. Comp. Psychol.*, 1941, **32**, 417-473.
22. Vaughn, C. L. Factors in rat learning. *Comp. Psychol. Monog.*, 1937, **14**.
23. Washburn, M. F. Hunger and speed of running as factors in maze-learning in mice. *J. Comp. Psychol.*, 1926, **6**, 181-187.
24. Zubin, J. Socio-biological types and methods for their isolation. *Psychiatry*, 1938, **1**, 237-247.